Robust energy-efficient MIMO transmission for cognitive vehicular networks

Article (Accepted Version)

Robust Energy-Efficient MIMO Transmission for Cognitive Vehicular Networks

Daxin Tian, *Senior Member, IEEE,* Jianshan Zhou, Zhengguo Sheng, and Victor C.M. Leung, *Fellow, IEEE,*

***Abstract*—This work investigates a robust energy-efficient solution for multiple-input-multiple-output (MIMO) transmissions in cognitive vehicular networks. Our goal is to design an optimal MIMO beamforming for secondary users (SUs) considering imperfect interference channel state information (CSI). Specifically, we optimize the energy efficiency (EE) of SUs, given that the transmission power constraint, the robust interference power constraint and the minimum transmission rate are satisfied. To solve the optimization problem, we first characterize the uncertainty of CSI by bounding it in a Frobenius-norm-based region and then equivalently convert the robust interference constraint to a linear matrix inequality. Furthermore, a feasible ascent direction approach is proposed to reduce the optimization problem into a sequential linearly constrained semi-definite program, which leads to a distributed iterative optimization algorithm for deriving the robust and optimal beamforming. The feasibility and convergence of the proposed algorithm is theoretically validated, and the final experimental results are also supplemented to show the strength of the proposed algorithm over some conventional schemes in terms of the achieved EE performance and robustness.**

***Index Terms*—Vehicular communications, cognitive radio, MIMO transmissions, energy efficiency**

I. Introduction

Advanced vehicular communication and networking technology plays an important role in the design of wireless connected vehicles and shows great potential to accelerate full deployment of Intelligent Transportation Systems (ITS) [1]–[3]. However, the dramatically increasing demand of vehicular telematics applications and infotainment services (e.g., HD movies and music online, data transfer, web browsing, and etc.), which highly relies on wireless communications, makes radio spectrum a very scarce and precious resource [4]–[6]. To make the best use of such resource, the so called cognitive vehicular networks, which combines the cognitive radio technology (CR) with vehicular communications, can share radio spectrum in an efficient and flexible way. In cognitive vehicular networks, licensed roadside communication infrastructure (e.g., cellular base station and wireless access points) or licensed vehicles can be regarded as primary users (PUs), while other unlicensed vehicular terminals are typically treated as secondary users (SUs) [7]. Meanwhile, the network consisting of terminals equipped with multiple antennas is referred to as a multiple-input multiple-output (MIMO) system, which is considered to be an emerging technology that can exploit spatial diversity [8], [9].

It sees that the green radio technologies have become an inevitable and critical trend in wireless communication networks [10]. Particularly, energy-efficient performance is a key concern in optimal designs of vehicular techniques deployed in V2I scenarios (e.g., downlink traffic scheduling [11], spectrum access [12], cooperative scheme [13], etc.). As argued by [10], financial consideration relevant to energy costs of base stations, roadside wireless nodes or some electronic vehicles, and environmental concerns resulting from power consumption are major factors to motivate green radio communications. The objective of green communications is to improve energy efficiency of radio systems, which include traditional cellular networks and vehicular networks. In addition, vehicle electrification has been recognized as an emerging trend induced by the forthcoming smart grid [14]. With the penetration of purely electric-powered or hybrid electric vehicles, energy saving is expected to be a much more essential issue in connected vehicle technologies in near future. Therefore, in this work, we focus on energy efficiency (EE) performance, which is characterized by the maximum amount of mutual information delivered by consuming per unit joule [15]

Generally, since the co-existing nature, the total interferences from SUs to a PU should be always restricted and required to be below a tolerable level [16]. There are some other challenges to obtain an optimal MIMO transmission design with the aforementioned considerations for a CR-enabled vehicular network: i) the MIMO beamforming design for a CR-enabled vehicular network is indeed a typical non-convex optimization problem; and ii) many real-life factors such as feed-back channel noises, time latencies or frequency offsets in SU-to-PU channels, will result in imperfect channel state information (CSI) obtained by SUs. More importantly, some characteristics of vehicular scenarios play a significant and unique role in design of vehicular MIMO transmission equipped with CR technology, which differentiate a CR-enabled network from conventional stationary/mobile CR networks. For example, the impact of vehicular mobility as well as the pattern of traffic flows on roads is recognized as the main factor characterizing a CR-enabled vehicular network. The vehicular mobility can lead to the dynamically changing spectrum availability perceived by CR-enabled vehicles, which

This research was supported in part by the National Natural Science Foundation of China under Grant nos. U1564212, 61103098, 91118008.

D. Tian and J. Zhou are with Beijing Key Laboratory for Cooperative Vehicle Infrastructure Systems & Safety Control, School of Transportation Science and Engineering, Beihang University, Beijing 100191, China (e-mail: tiandaxin@gmail.com).

Z. Sheng is with Department of Engineering and Design, the University of Sussex, Richmond 3A09, UK.

V. Leung is with Department of Electrical and Computer Engineering, The University of British Columbia, Vancouver, B.C., V6T 1Z4 Canada.

in turn makes it more challenging to obtain accurate sensing information on the cognitive-to-primary propagation channels. The vehicular motion may also cause the time-variation of the propagation channels, impacting on the correctness of the estimation of the secondary vehicle-to-primary vehicle channels. Besides, absence or impracticality of centralized computation indeed increases the complexity of MIMO transmission design in vehicular scenarios. Therefore, an appropriate MIMO beamforming design solution needs to capture inaccuracies in the information of SU-to-PU vehicular channels, to be robust and suitable for on-line distributed implementation so that it can allow to some extent accommodation of dynamic nature of propagation channels in vehicular scenarios.

Up to now, there have been many studies focusing on EE transmission of MIMO systems such as [17]–[21]. The authors in [17] aimed to achieve a high energy efficiency and proposed a EE strategy for V2I and I2I (infrastructure-to-infrastructure) communications based on cooperative relay and cooperative MIMO transmissions. Although their method was proven to be more effective than traditional SISO transmission strategy, it did not consider some real-life requirements such as the maximum power budget and the required data rate. In [18], the researchers considered to obtain an optimal trade-off between the sum of data rates and the power cost for MIMO broadcast channels. But, the effect of interference power on the EE optimization was not involved in their study.

In our model, we adopt a specific mathematical formulation of the performance metric on EE, i.e., the ratio of the data rate over the power consumed to perform MIMO transmissions, which is similar to the models proposed in [19]–[21]. In these models, only the conditions on maximum power budget and tolerable interference power were considered, and they are under the assumption of perfect CSI. But, CSI is usually imperfect to SUs in reality, and PUs may not be aware of the existence of some secondary transmission links, such that the secondary transmitters should be in charge of restricting their interference power to the primary receivers. There are also some researches focusing on the robust optimization for MIMO transmissions in CR networks [22]–[26]. The work of [22] presented an iterative robust beamforming algorithm to maximize the received signal-to-interference-plus-noise ratios (SINRs) in the worst situation. This approach belongs to a type of conservative decision making and does not aim at EE optimization. In [23], the authors formulated the uncertainty of imperfect CSI as a kind of added Gaussian noise, which was a probabilistic model. These models mainly focus on maximizing the service probability of SUs rather than the energy efficiency. In [24], the payoff function of every player (i.e., every SU) was modeled as its maximum Shannon mutual information, which was subject to the power constraint. Similarly, the payoff function of [25] was also represented as the maximum achievable rate on an individual SU link. In [26], a robust beamforming optimization was proposed to minimize the overall mean-square error in MIMO transmissions, which was based on a cyclic block coordinate ascent algorithm. The transmission power consumption and the interference power constraints set up in [26] are similar to those in our work. The difference is that we consider to satisfy the minimum transmission rate required by SUs. Additionally, it is worth pointing out that the studies presented in [22], [25], [26] have shown the strength of the S-Procedure theorem [27] to convert the robust interference constraint to a linear matrix inequality. This idea behind the S-Procedure also motivates us to perform the constraint transformation so as to make the EE optimization tractable. The essential difference between our work and those aforementioned lies in the optimization objective and the constraint sets it is subject to and in the idea behind the optimal beamforming design.

We aim at optimizing EE MIMO transmission for CR-enabled vehicular networks by modeling EE performance metric as the optimization objective function. For the optimization model, we present a robust and distributed solution which takes into account the uncertainty of the interference channel state. The proposed algorithm is suitable to be implemented in a distributed and online manner, which allows for dynamic nature of propagation channels in vehicular scenarios. The following summarize our main contributions of the paper:

i) We jointly consider three types of constraints in our EE MIMO transmission optimization model simultaneously, including *the maximum energy budget constraint*, *the minimum data rate constraint*, and *the robust interference constraint*, such that we can derive an appropriate joint power allocation and beamforming pattern for each SU pair.

ii) Considering the imperfect CSI, we mathematically formulate the robust interference constraint to make the model tractable, such that we can provide a robust design. Specifically, an elliptical uncertainty region represented by Frobenius norm is adopted to characterize the inaccuracy of SU-to-PU channel state. Then, by exploiting the well-known S-Procedure [27], we perform an equivalent transformation on the robust interference constraint and re-express it as a linear matrix inequality (LMI), which is more convenient to be solved by using linear semi-definite programming (SDP) techniques.

iii) We propose a feasible ascent direction method based on sequential linear programming. We obtain a feasible ascent direction by solving the modified objective subject to the linearizion of the minimum data rate constraint and to the power consumption constraint combined with the LMI. The sequential linearly constrained programming sub-problem can be achieved effectively since its objective and constraint functions are all linearized, which belongs to a typical linear semi-definite program. Based on the feasible direction obtained previously, we transform the original EE optimization to a linear searching sub-problem which can be effectively solved to derive an optimal iteration step to generate a new iterator. Hence, we yield a robust iterative optimization solution for EE MIMO transmissions. Furthermore, we also theoretically provide the feasibility and convergence of sequential iterators generated at each SU and show the strength of our proposed algorithm through experiments.

The remainder of this paper is organized as follows. We introduce preliminaries of the basic optimization model as well as its modified form with consideration of imperfect channel state information in Section II. Section III elaborates on our robust distributed iterative optimization solution. In Section IV we theoretically analyze the feasibility and convergence of

TABLE I
SUMMARY OF BASIC MATHEMATICAL SYMBOLS AND OPERATIONS

Symbols&Operations	Meaning
$\mathbb{R}$	space of real numbers
$\mathbb{C}$	space of complex numbers
$\mathbb{M}_+$	space of Hermitian positive semi-definite matrices
$\Re(\mathbf{X})$	real part of a complex matrix $\mathbf{X}$
$\Im(\mathbf{X})$	complex part of a complex matrix $\mathbf{X}$
$\mathbf{X}^T$	transpose operation on $\mathbf{X}$
$\mathbf{X}^\dagger$	conjugate transpose operation on $\mathbf{X}$
$\mathbf{X}^{-1}$	inverse operation on $\mathbf{X}$
$\mathbf{X}^*$	component-wise complex conjugation of $\mathbf{X}$
$\mathrm{vec}(\mathbf{X})$	stacking vectorization of $\mathbf{X}$
$\mathrm{Tr}(\mathbf{X})$	trace of $\mathbf{X}$
$\det(\mathbf{X})$	determinant of $\mathbf{X}$
$\otimes$	Kronecker product operation
$\mathbb{E}(\cdot)$	expectation of a variable
$\mathbf{X} \succ \mathbf{0}$ ($\mathbf{X} \succeq \mathbf{0}$)	$\mathbf{X}$ is positive definite (semidefinite)
$\mathbf{I}_n$	the $n \times n$ identity matrix
$\mathbf{0}_n$	the $n \times n$ zero matrix

iterators generated by solving sequential linearly constrained semi-definite programming sub-problems. Section V presents the experiment results, and Section VI concludes this work. In addition, we summarize basic mathematical symbols and operations as well as their meaning in Table I. Unless otherwise specified, the vectors mentioned in this work are column vectors and denoted by boldface lowercase letters. Throughout this paper the boldface uppercase letters are used to denote the matrices.

II. PRELIMINARIES

A. Basic Optimization Model

Without loss of generality, we consider a MIMO CR-enabled vehicular network shown in Fig.1, which consists of a group of unlicensed vehicles and a licensed vehicle. These vehicular communication terminals are equipped with multiple antennas. As illustrated in Fig.1, the V2I communication constructs a primary network, in which an infrastructure node (e.g., a base station) treated as a primary transmitter (named Primary Tx) serves this licensed vehicle who is a primary receiver (named Primary Rx). Other unlicensed vehicles sharing the same frequency channel with the primary receiver are regarded as some secondary users with V2V communications, whose links form a secondary network. In particular, there exist interference links between the secondary transmitters (named Secondary Tx) and the primary receiver and their undesired secondary receivers (named Secondary Rx).

In general, a group of vehicles in a steady traffic flow move with the same steady speed (i.e., the average speed of this traffic flow), such that the relative geographic distances within the vehicular group, i.e., the vehicular topology structure, are steady in a certain duration as well. Hence, for the steady traffic flow scenario, it is suitable to assume that the propagation channels within the vehicular group in a CR-enabled network are stable (at least, during a certain

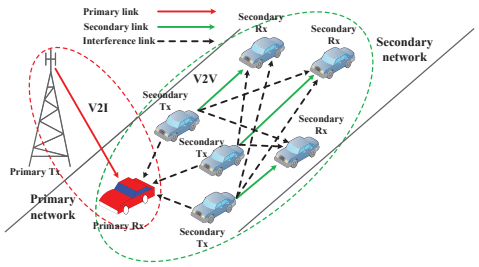


Fig. 1. The system model for MIMO CR-enabled vehicular networks.

period of time). At this point, we can model the vehicular propagation channels with time-invariant channel matrices. Let the number of secondary links in the CR-enabled vehicular network considered under Gaussian-interference channel be J. Each secondary link is associated with a transmitter-receiver pair $(Tx_j, Rx_j) \in \{(Tx_j, Rx_j)\,|j \in \mathscr{J}\}$. $\mathscr{J}$ denotes the set of those transmitter-receiver links whose cardinality is J, i.e., $\mathscr{J} := \{1, 2, \ldots, J\}$, $J = |\mathscr{J}|$. The transmitter Tx_j at the j-th link is equipped with m_j antennas, while the corresponding receiver Rx_j has n_j antennas and $m_j \leq n_j$. The transmitter Tx_j transmits a complex signal vector $\mathbf{s}_j$ of dimension m_j, i.e., $\mathbf{s}_j \in \mathbb{C}^{m_j \times 1}$ such that a complex baseband signal vector of dimension n_j, $\mathbf{r}_j \in \mathbb{C}^{n_j \times 1}$, can be received at the receiver Rx_j. Let $\mathbf{z}_j \in \mathbb{C}^{n_j \times 1}$ denote the noise floor plus interference from primary users on the same channel which can be modeled as a zero-mean circularly symmetric complex Gaussian noise vector with a nonsingular covariance matrix $\mathbf{R}_j \in \mathbb{C}^{n_j \times n_j}$, i.e., $\mathbf{z}_j \sim N(0, \mathbf{R}_j)$, $\mathbb{E}(\mathbf{z}_j \mathbf{z}_j{}^\dagger) = \mathbf{R}_j$. Thus, with $\mathbf{H}_{j,i} \in \mathbb{C}^{n_j \times m_i}$ representing the complex channel matrix between the transmitter Tx_i and the receiver Rx_j, we can formulate the received signal at the receiver Rx_j, $\mathbf{r}_j$, as

$$\mathbf{r}_j = \sum_{i \in \mathscr{J}} \mathbf{H}_{j,i}\mathbf{s}_i + \mathbf{z}_j = \mathbf{H}_{j,j}\mathbf{s}_j + \sum_{i \in \mathscr{J}_{-j}} \mathbf{H}_{j,i}\mathbf{s}_i + \mathbf{z}_j \quad (1)$$

where $\mathscr{J}_{-j} := \{i | i \in \mathscr{J}, i \neq j\}$ and for all $j \in \mathscr{J}$ the channel matrix $\mathbf{H}_{j,j}$ can be considered to be nonzero.

Furthermore, we also consider that the number of the original independent data substreams transmitted in the j-th link is l_j ($l_j \leq m_j$). These original substreams, denoted by $\mathbf{x}_j$ of dimension l_j, i.e., $\mathbf{x}_j \in \mathbb{C}^{l_j \times 1}$, are mapped to $\mathbf{s}_j$ through being multiplied by an precoding matrix $\mathbf{M}_j \in \mathbb{C}^{m_j \times l_j}$. Here this matrix $\mathbf{M}_j$ in fact represents a certain precoding or transmit-beamforming strategy for alleviating self-interference of the transmitter Tx_j. That is, the signal vector $\mathbf{s}_j$ in the j-th link can be rewritten as

$$\mathbf{s}_j = \mathbf{M}_j \mathbf{x}_j \quad (2)$$

At the associated receiver Rx_j of the j-th link, the received signal vector $\mathbf{r}_j$ can be mapped to l_j independent data substreams by a decoding operation. That is, Rx_j transforms $\mathbf{r}_j$ to another complex signal vector of dimension l_j, $\mathbf{y}_j \in \mathbb{C}^{l_j \times 1}$, by using a deconding matrix $\mathbf{D}_j \in \mathbb{C}^{l_j \times n_j}$ as following

$$\mathbf{y}_j = \mathbf{D}_j \mathbf{r}_j \quad (3)$$

Substituting Equations (2) and (3) into (1), we can get

$$\mathbf{y}_j = \mathbf{D}_j\mathbf{H}_{j,j}\mathbf{M}_j\mathbf{x}_j + \sum_{i\in\mathscr{J}_{-j}} \mathbf{D}_j\mathbf{H}_{j,i}\mathbf{M}_i\mathbf{x}_i + \mathbf{D}_j\mathbf{z}_j \quad (4)$$

Note that the matrix $\mathbf{M}_j$ characterizes the associated transmitted signal vector $\mathbf{s}_j$ in terms of transmission power consumption. Following some typical work such as [19], [19], [21], [25], [26], we represent the total transmission power consumption of the j-th link in the MIMO CR network as $p_j = \beta\mathrm{Tr}(\mathbf{Q}_j) + p_{j,c}$ where $\mathbf{Q}_j$ denotes the covariance matrix corresponding to $\mathbf{M}_j$, i.e., $\mathbf{Q}_j := \mathbb{E}\left(\mathbf{M}_j{\mathbf{M}_j}^\dagger\right)$, and is Hermitian positive semidefinite, i.e., $\mathbf{Q}_j \in \mathbb{M}_+^{m_j\times m_j}$. The parameter $p_{j,c}$ is used to denote the total circuit-related power consumption at both Tx_j and Rx_j. β is a positive coefficient ($\beta \in (0,1)$), whose reciprocal indicates efficiency of the power amplifier. Since a SU's transmission power is limited, introducing an individual power threshold $p_j^{\max} \in \mathbb{R}$ we can formulate an inequality constraint with respect to $\mathbf{Q}_j$

$$p_j = \beta\mathrm{Tr}(\mathbf{Q}_j) + p_{j,c} \le p_j^{\max} \quad (5)$$

for $\forall j \in \mathscr{J}$.

To proceed the model, we introduce a complex matrix, $\mathbf{U}_j \in \mathbb{C}^{u\times m_j}$, to represent the channel gain between a primary receiver equipped with u antennas and the j-th secondary user in the MIMO CR network. Thus, we consider to restrict the instantaneous interference resulting from the j-th secondary user to this primary receiver in the CR network where primary and secondary links coexist. At this point, we have the following inequality constraint on the channel interference

$$\sum_{j\in\mathscr{J}} \gamma_j := \sum_{j\in\mathscr{J}} \mathrm{Tr}\left(\mathbf{U}_j\mathbf{Q}_j{\mathbf{U}_j}^\dagger\right) \le \gamma^{\max} \quad (6)$$

where $\gamma^{\max} \in \mathbb{R}$ denotes a predefined maximum interference allowed at the primary user. It should be noted that the information on $\mathbf{U}_i$ for $\forall i \in \mathscr{J}$ has to be fed back to the individual j-th receiver Rx_j when adopting (6) in an optimization scheme. However, as suggested by many existing literature [24], [26], [28], it is appropriate to pre-partition the total power interference that a primary user can tolerate in per-secondary link portions, which can be helpful to realize a distributed implementation with low-complexity and to guarantee possible quality-of-service requirements, while some other study cases [20], [21], [29] have provided their beamforming solutions where the upper bound of a total aggregate interference is not divided. For example, in [30], the aggregate interference constraint directly is combined with the power constraint via introducing auxiliary variables. Nevertheless, certain complexity may be increased when considering performing these solutions in distributed scenarios, since additional efforts (e.g., decomposition techniques) are needed to dynamically allocate the total power-interference to CR-enabled secondary links. Based on the considerations above, we adopt the formation of the per-cognitive radio link constraint in our model, which is more stringent than the total-power interference constraint. That is, to make the computation more practical, we can pre-divide such a total instantaneous interference $\gamma^{\max}$ that the primary user can tolerate into a series of interference thresholds associated with each individual secondary link, i.e., $\left\{\gamma_j^{\max} | \gamma_j^{\max} > 0, \gamma^{\max} = \sum_{j\in\mathscr{J}} \gamma_j^{\max}\right\}$, such that

$$\gamma_j = \mathrm{Tr}\left(\mathbf{U}_j\mathbf{Q}_j{\mathbf{U}_j}^\dagger\right) \le \gamma_j^{\max} \quad (7)$$

for $\forall j \in \mathscr{J}$. It is worth pointing out that each individual interference threshold, $\gamma_j^{\max} \in \mathbb{R}$, can be pre-specified, which may depend on some certain QoS demands of the corresponding individual transmitter-receiver pair (Tx_j, Rx_j). Hence we can use the inequality constraint (7) in our following optimization formulation instead of the aggregative form (6).

Based on (4), the j-th link's theoretical maximum information rate r_j can be defined as

$$\begin{aligned} r_j =& \log_2\det\left(\mathbf{\Phi}(\mathbf{D}_j, \mathscr{Q}_{-j}) + \mathbf{D}_j\mathbf{H}_{j,j}\mathbf{Q}_j\mathbf{H}_{j,j}^\dagger\mathbf{D}_j^\dagger\right) \\ &- \log_2\det\left(\mathbf{\Phi}(\mathbf{D}_j, \mathscr{Q}_{-j})\right) \end{aligned} \quad (8)$$

where $\mathscr{Q}_{-j}$ denotes the collection of the covariance matrices of the secondary transmitter-receiver pairs except for the j-th one $\mathbf{Q}_j$, i.e., $\mathscr{Q}_{-j} := \{\mathbf{Q}_i | \forall i \in \mathscr{J}_{-j}\}$. $\mathbf{\Phi}(\mathbf{D}_j, \mathscr{Q}_{-j})$ is defined as the multiple-user interference plus noise observed at the j-th secondary user

$$\mathbf{\Phi}(\mathbf{D}_j, \mathscr{Q}_{-j}) := \sum_{i\in\mathscr{J}_{-j}} \mathbf{D}_j\mathbf{H}_{j,i}\mathbf{Q}_i\mathbf{H}_{j,i}^\dagger{\mathbf{D}_j}^\dagger + \widetilde{\mathbf{R}}_j \quad (9)$$

where $\widetilde{\mathbf{R}}_j$ is a positive semidefinite covariance matrix of $l_j\times l_j$ dimension whose radius is the same as that of $\mathbf{R}_j$, i.e., $\widetilde{\mathbf{R}}_j \in \mathbb{C}^{l_j\times l_j}$ and $\rho\left(\widetilde{\mathbf{R}}_j\right) = \rho\left(\mathbf{R}_j\right)$.

Generally, r_j should not be too low so as to guarantee a QoS transmission required at the j-th secondary user's link. Namely, when r_j could not satisfy a required minimum transmission rate, the corresponding secondary transmission link should be closed. Otherwise, it would lead to unnecessary interference or energy consumption. Therefore, we introduce a transmission rate constraint

$$r_j \ge r_j^{\min} \quad (10)$$

for $\forall j \in \mathscr{J}$ where $r_j^{\min} \in \mathbb{R}$ denotes a pre-specified transmission rate threshold.

By combining the inequality constraints (5), (7) and (10), a basic model formulating the energy-efficient transmission optimization problem of the overall MIMO CR network can be expressed as

$$\underset{\{(\mathbf{Q}_j,\mathbf{D}_j)|\forall j\in\mathscr{J}\}}{\text{maximize}} : \sum_{j\in\mathscr{J}} f_j\left(\mathbf{Q}_j, \mathbf{D}_j \mid \mathscr{H}_j, \mathscr{Q}_{-j}\right) := \sum_{j\in\mathscr{J}} \frac{r_j}{p_j} \quad (11a)$$

$$s.t.\ p_j \le p_j^{\max} \quad (11b)$$

$$\gamma_j \le \gamma_j^{\max} \quad (11c)$$

$$r_j \ge r_j^{\min} \quad (11d)$$

$$\mathbf{Q}_j \in \mathbb{M}_+^{m_j\times m_j} \quad (11e)$$

$$\mathbf{D}_j \in \mathbb{C}^{l_j\times n_j}, j \in \mathscr{J} \quad (11f)$$

where the set $\mathscr{H}_j$ (for $\forall j \in \mathscr{J}$) is defined as the collection of all the channel matrix related to the j-th receiver Rx_j, i.e., $\mathscr{H}_j := \{\mathbf{H}_{j,i} | \mathbf{H}_{j,i} \in \mathbb{C}^{n_j\times m_i}, \forall i \in \mathscr{J}\}$. $f_j\left(\mathbf{Q}_j, \mathbf{D}_j \mid \mathscr{H}_j, \mathscr{Q}_{-j}\right)$ is used to denote r_j/p_j, i.e.,

$f_j\left(\mathbf{Q}_j, \mathbf{D}_j \mid \mathscr{H}_j, \mathscr{Q}_{-j}\right) := r_j/p_j$. According to the optimization objective function (11a) given in the basic model (11), we point out that as a energy efficiency metric $f_j\left(\mathbf{Q}_j, \mathbf{D}_j \mid \mathscr{H}_j, \mathscr{Q}_{-j}\right)$ actually indicates the quantity of information successfully transmitted at the secondary link per unit of energy consumed.

B. Optimization Formulation with SU-to-PU Channel Uncertainty

In an actual MIMO CR communication scenario, the channel information between the secondary and the primary users is usually imperfect. Thus, a robust formulation should take into consideration of a certain error (uncertainty), $\Delta\mathbf{U}_j$, existing in the SU-to-PU channel matrix $\mathbf{U}_j$. Representing the estimated channel matrix as $\widetilde{\mathbf{U}}_j$, we can re-formulate the observed channel matrix as $\mathbf{U}_j = \widetilde{\mathbf{U}}_j + \Delta\mathbf{U}_j$. Following the works [22], [25], [26], we assume that the error matrix $\Delta\mathbf{U}_j$ (for $\forall j \in \mathscr{J}$) is bounded. That is, it should be limited to a finite set

$$\mathscr{U}_j(\sigma_j) := \left\{\Delta\mathbf{U}_j | \Delta\mathbf{U}_j \in \mathbb{C}^{u\times m_j}, \mathrm{Tr}\Big(\Delta\mathbf{U}_j\Delta\mathbf{U}_j^{\dagger}\Big) \le \sigma_j^2\right\} \tag{12}$$

where the parameter $\sigma_j \in \mathbb{R}$ ($\sigma_j > 0$) implies the radius of $\Delta\mathbf{U}_j$, which can quantify the error level of the channel or the degree of the channel uncertainty. Consequently, accounting for $\Delta\mathbf{U}_j$ we get the following robust optimization formulation

$$\underset{\{(\mathbf{Q}_j,\mathbf{D}_j)|\forall j\in\mathscr{J}\}}{\text{maximize}} : \sum_{j\in\mathscr{J}} f_j\left(\mathbf{Q}_j, \mathbf{D}_j \mid \mathscr{H}_j, \mathscr{Q}_{-j}\right) := \sum_{j\in\mathscr{J}} \frac{r_j}{p_j} \tag{13a}$$

$$s.t.\ p_j \le p_j^{\max} \tag{13b}$$

$$\gamma_j \le \gamma_j^{\max}, \forall\Delta\mathbf{U}_j \in \mathscr{U}_j(\sigma_j) \tag{13c}$$

$$r_j \ge r_j^{\min} \tag{13d}$$

$$\mathbf{Q}_j \in \mathbb{M}_+^{m_j\times m_j} \tag{13e}$$

$$\mathbf{D}_j \in \mathbb{C}^{l_j\times n_j}, j \in \mathscr{J} \tag{13f}$$

III. Robust Distributed Optimization Algorithm

In fact, it is impractical to deal with the optimization problem represented by Equations (13a)∼(13e) since it requires the global knowledge of the MIMO CR network and a centralized control or infrastructure to solve all the individual secondary users' decisions $\{(\mathbf{Q}_j, \mathbf{D}_j)\,|\forall j \in \mathscr{J}\}$. To induce a robust distributed optimization algorithm, we can fix other secondary users' decision variables except for the j-th one's, $\{(\mathbf{Q}_i, \mathbf{D}_i)\,|\forall i \in \mathscr{J}_{-j}\}$, for the j-th secondary user. Namely, in an energy-efficient optimization programming associated with the j-th secondary transmission pair (Tx_j, Rx_j) $(\mathbf{Q}_j, \mathbf{D}_j)$ are treated as the decision variables while others $\{(\mathbf{Q}_i, \mathbf{D}_i)\,|\forall i \in \mathscr{J}_{-j}\}$ are kept as parameters. Thus, such an optimization programming can be locally solved at the individual secondary user (Tx_j, Rx_j). Accordingly, the individual energy-efficient transmission optimization problem of the j-th secondary user can be expressed as

$$\underset{\{(\mathbf{Q}_j,\mathbf{D}_j)\}}{\text{maximize}} : g_j\left(\mathbf{Q}_j, \mathbf{D}_j \mid \mathscr{H}_j, \mathscr{Q}_{-j}\right) \tag{14a}$$

$$s.t.\ p_j \le p_j^{\max} \tag{14b}$$

$$\gamma_j \le \gamma_j^{\max}, \forall\Delta\mathbf{U}_j \in \mathscr{U}_j(\sigma_j) \tag{14c}$$

$$r_j \ge r_j^{\min}, j \in \mathscr{J} \tag{14d}$$

$$\mathbf{Q}_j \in \mathbb{M}_+^{m_j\times m_j}, \mathbf{D}_j \in \mathbb{C}^{l_j\times n_j} \tag{14e}$$

where we define

$$\begin{aligned} g_j\left(\mathbf{Q}_j, \mathbf{D}_j \mid \mathscr{H}_j, \mathscr{Q}_{-j}\right) :=& f_j\left(\mathbf{Q}_j, \mathbf{D}_j \mid \mathscr{H}_j, \mathscr{Q}_{-j}\right) \\ &+ \mathcal{F}_j\left(\mathbf{Q}_j \mid \mathscr{H}_j, \mathscr{Q}_{-j}\right). \end{aligned} \tag{15}$$

The function $\mathcal{F}_j\left(\mathbf{Q}_j \mid \mathscr{H}_j, \mathscr{Q}_{-j}\right)$ is the sum of the energy efficiency metrics, $\{f_i\left(\mathbf{Q}_i, \mathbf{D}_i \mid \mathscr{H}_i, \mathscr{Q}_{-i}\right)|\forall i \in \mathscr{J}_{-j}\}$, except for the j-th one, i.e., $\mathcal{F}_j\left(\mathbf{Q}_j \mid \mathscr{H}_j, \mathscr{Q}_{-j}\right) = \sum_{i\in\mathscr{J}_{-j}} f_i\left(\mathbf{Q}_i, \mathbf{D}_i \mid \mathscr{H}_i, \mathscr{Q}_{-i}\right)$. Now, we first transform the robust channel interference constraint in (14c) to another form, which is more useful for solving the energy-efficient transmission optimization problem. Specifically, given the bound set of $\mathbf{U}_j$ with the positive parameter σ_j, $\mathscr{U}_j\left(\sigma_j\right)$ (see (12)), we have the following corollary:

Corollary 1. *Consider $\mathbf{U}_j = \widetilde{\mathbf{U}}_j + \Delta\mathbf{U}_j$ and the bound set $\mathscr{U}_j\left(\sigma_j\right)$ in (12) so that $\Delta\mathbf{U}_j \in \mathscr{U}_j\left(\sigma_j\right)$. The inequality constraint (14c) is equivalent to*

$$\begin{aligned} &\left(\mathrm{vec}\left(\Delta\mathbf{U}_j^{\dagger}\right)\right)^{\dagger}\left(-\mathbf{I}_u \otimes \mathbf{Q}_j\right)\mathrm{vec}\left(\Delta\mathbf{U}_j^{\dagger}\right) \\ &+ 2\Re\left(-\mathrm{vec}\left(\mathbf{Q}_j^{\dagger}\widetilde{\mathbf{U}}_j^{\dagger}\right)^{\dagger}\mathrm{vec}\left(\Delta\mathbf{U}_j^{\dagger}\right)\right) \\ &+ \gamma_j^{\max} - \mathrm{Tr}\left(\widetilde{\mathbf{U}}_j\mathbf{Q}_j\widetilde{\mathbf{U}}_j^{\dagger}\right) \ge 0,\ \left\|\mathrm{vec}\left(\Delta\mathbf{U}_j^{\dagger}\right)\right\|_2 \le \sigma_j \end{aligned} \tag{16}$$

for $\forall j \in \mathscr{J}$ where $\|\cdot\|_2$ denotes the Euclidean norm (i.e., 2-norm operator). (16) can be further transformed to a linear matrix inequality constraint:

$$\begin{bmatrix} \lambda_j\mathbf{I}_{m_j u} - \left(\mathbf{I}_u \otimes \mathbf{Q}_j\right) & -\mathrm{vec}\left(\mathbf{Q}_j^{\dagger}\widetilde{\mathbf{U}}_j^{\dagger}\right) \\ -\left(\mathrm{vec}\left(\mathbf{Q}_j^{\dagger}\widetilde{\mathbf{U}}_j^{\dagger}\right)\right)^{\dagger} & \gamma_j^{\max} - \mathrm{Tr}\left(\widetilde{\mathbf{U}}_j\mathbf{Q}_j\widetilde{\mathbf{U}}_j^{\dagger}\right) - \lambda_j\sigma_j^2 \end{bmatrix} \succeq \mathbf{0} \tag{17}$$

where $\lambda_j \ge 0$ is a non-negative real number.

Proof: See the Appendix A-C. ∎

Based on Corollary 1, we then equivalently reshape the optimization model using the linear matrix inequality (17) instead of (14b). Namely, we have

$$\underset{\{(\mathbf{Q}_j,\mathbf{D}_j,\lambda_j)\}}{\text{maximize}} : g_j\left(\mathbf{Q}_j, \mathbf{D}_j \mid \mathscr{H}_j, \mathscr{Q}_{-j}\right) \tag{18a}$$

$$s.t.\ \left(\mathbf{Q}_j, \mathbf{D}_j, \lambda_j\right) \in \mathscr{S}_j \tag{18b}$$

where the set $\mathscr{S}_j$ collects the constraints on $(\mathbf{Q}_j, \mathbf{D}_j, \lambda_j)$, denoted by

$$\mathscr{S}_j := \left\{(\mathbf{Q}_j, \mathbf{D}_j, \lambda_j) \mid (14b), (14d), (14e) \text{ and } (17)\right\}. \tag{19}$$

Note that the objective function indicating the energy efficiency as given in (18a) is not concave or convex in $(\mathbf{Q}_j, \mathbf{D}_j)$. It is difficult to directly solve a global optimum point of this

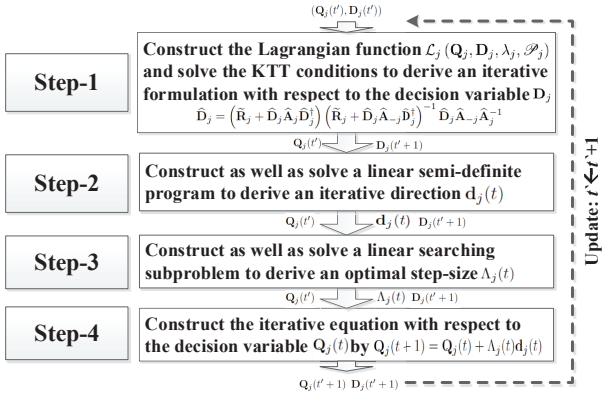


Fig. 2. The key steps of the proposed iterative approach.

problem as shown in (18a)∼(18b). Alternatively, we would like to build a iterative algorithm to approach an optimum $\left(\widehat{\mathbf{Q}}_j, \widehat{\mathbf{D}}_j\right)$.

Our proposed approach is illustrated in Fig. 2, which involves four key operation steps: (i) a Lagrangian function as well as the associated KTT conditions are established, whose solution can lead to a useful iterative formulation to approach the optimal $\widehat{\mathbf{D}}_j$; then, (ii) a linear semi-definite programming model is further developed to derive an feasible iterative direction; (iii) on the basis of such a direction, a one-dimension searching optimization is solved to determine an optimal iteration step-size; finally, (iv) a new iterator approaching to $\widehat{\mathbf{Q}}_j$ can be yielded through computing a simple incremental equation that lumps the current point and the searching direction and the optimal searching step-size. The mathematical presentations corresponding to each operation step in the proposed iteration framework shown in Figure 2 are detailed in the following subsections, respectively.

A. Karush-Kuhn-Tucker (KKT) conditions

According to (18), we establish the Lagrangian function with respect to $(\mathbf{Q}_j, \mathbf{D}_j)$ and λ_j as

$$\begin{aligned}\mathcal{L}_j\left(\mathbf{Q}_j, \mathbf{D}_j, \lambda_j, \mathscr{P}_j\right) := & -g_j\left(\mathbf{Q}_j, \mathbf{D}_j \mid \mathscr{H}_j, \mathscr{Q}_{-j}\right) \\ & + \alpha_{j,1}\left(p_j - p_j^{\max}\right) \\ & + \alpha_{j,2}\left(\gamma_j - \gamma_j^{\max}\right) \\ & + \sum_{i \in \mathscr{J}} \alpha_{j,i,3}\left(r_i^{\min} - r_i\right)\end{aligned} \tag{20}$$

where $\mathscr{P}_j$ is the set of non-negative Lagrangian multipliers, i.e., $\mathscr{P}_j := \left\{\alpha_{j,1} \geq 0, \alpha_{j,2} \geq 0, \{\alpha_{j,i,3} \geq 0\}_{\forall i \in \mathscr{J}}\right\}$.

Note that the decision variable, $\mathbf{D}_j$, only explicitly involved in the terms $f_j\left(\mathbf{Q}_j, \mathbf{D}_j \mid \mathscr{H}_j, \mathscr{Q}_{-j}\right)$ and r_j, is not coupled with the uncertain channel interference $\mathbf{U}_j$. We turn to focus on the Karush-Kuhn-Tucker (KKT) condition of (18) with respect to $\mathbf{D}_j$, which can be expressed as

$$\bigtriangledown_{\mathbf{D}_j} \mathcal{L}_j\left(\widehat{\mathbf{Q}}_j, \widehat{\mathbf{D}}_j, \widehat{\lambda}_j, \mathscr{P}_j\right) = \mathbf{0} \tag{21}$$

where $\bigtriangledown_x$ represents the gradient operator with respect to the variable x. Since $\mathbf{D}_j$ is a complex matrix, $\bigtriangledown_{\mathbf{D}_j} \mathcal{L}_j\left(\widehat{\mathbf{Q}}_j, \widehat{\mathbf{D}}_j, \widehat{\lambda}_j, \mathscr{P}_j\right)$ is actually the complex-valued partial derivative of $\mathcal{L}_j\left(\mathbf{Q}_j, \mathbf{D}_j, \lambda_j, \mathscr{P}_j\right)$. The formula for computing the complex-valued gradient of a scalar function can be found in [31], [32]. Accordingly, the complex-valued gradient of the Lagrangian function with respect to $\mathbf{D}_j$ is yielded as

$$\bigtriangledown_{\mathbf{D}_j} \mathcal{L}_j\left(\mathbf{Q}_j, \mathbf{D}_j, \lambda_j, \mathscr{P}_j\right) = -\left(\frac{1}{p_j} + \alpha_{j,j,3}\right) \bigtriangledown_{\mathbf{D}_j} r_j \tag{22}$$

Hence, (21) is equivalent to $\bigtriangledown_{\mathbf{D}_j} r_j\big|_{\left(\widehat{\mathbf{Q}}_j, \widehat{\mathbf{D}}_j\right)} = \mathbf{0}$ since $\left(\frac{1}{p_j} + \alpha_{j,j,3}\right) \neq 0$. Given $\left(\widehat{\mathbf{Q}}_j, \widehat{\mathbf{D}}_j\right)$ is a KKT point (i.e., a stationary point), solving this equation results in

$$\widehat{\mathbf{D}}_j = \left(\widetilde{\mathbf{R}}_j + \widehat{\mathbf{D}}_j \widehat{\mathbf{A}}_j \widehat{\mathbf{D}}_j^{\dagger}\right)\left(\widetilde{\mathbf{R}}_j + \widehat{\mathbf{D}}_j \widehat{\mathbf{A}}_{-j} \widehat{\mathbf{D}}_j^{\dagger}\right)^{-1} \widehat{\mathbf{D}}_j \widehat{\mathbf{A}}_{-j} \widehat{\mathbf{A}}_j^{-1} \tag{23}$$

where $\widehat{\mathbf{A}}_j$ is defined as $\widehat{\mathbf{A}}_j := \sum_{i \in \mathscr{J}} \mathbf{H}_{j,i} \widehat{\mathbf{Q}}_i \mathbf{H}_{j,i}^{\dagger} \in \mathbb{M}_+^{n_j \times n_j}$ while $\widehat{\mathbf{A}}_{-j} := \sum_{i \in \mathscr{J}_{-j}} \mathbf{H}_{j,i} \widehat{\mathbf{Q}}_i \mathbf{H}_{j,i}^{\dagger} \in \mathbb{M}_+^{n_j \times n_j}$. For simplicity, we also define $\mathbf{A}_j := \sum_{i \in \mathscr{J}} \mathbf{H}_{j,i} \mathbf{Q}_i \mathbf{H}_{j,i}^{\dagger} \in \mathbb{M}_+^{n_j \times n_j}$ and $\mathbf{A}_{-j} := \sum_{i \in \mathscr{J}_{-j}} \mathbf{H}_{j,i} \mathbf{Q}_i \mathbf{H}_{j,i}^{\dagger} \in \mathbb{M}_+^{n_j \times n_j}$.

When considering that the MIMO interference channel with the precoding matrix, $\mathbf{M}_j$, and the power consumption (i.e., $\mathbf{Q}_j$), are given and fixed, maximizing the energy-efficiency, $g_j\left(\mathbf{Q}_j, \mathbf{D}_j \mid \mathscr{H}_j, \mathscr{Q}_{-j}\right)$, can be achieved through designing an appropriate decoding matrix, $\mathbf{D}_j$, based on Equation (23) above. Although it is not easy or even impractical to solve (23) to obtain an optimum $\widehat{\mathbf{D}}_j$, (23) leads to the following iterative equation which can be used to approach $\widehat{\mathbf{D}}_j$

$$\mathbf{D}_j(t'+1) = \mathbf{W}_1(t')\mathbf{W}_2(t')\mathbf{W}_3(t') \tag{24}$$

where each of the terms at the right side of (24) is defined as

$$\begin{cases} \mathbf{W}_1(t') = \left(\widetilde{\mathbf{R}}_j + \mathbf{D}_j(t')\mathbf{A}_j(t')\mathbf{D}_j^{\dagger}(t')\right) \\ \mathbf{W}_2(t') = \left(\widetilde{\mathbf{R}}_j + \mathbf{D}_j(t')\mathbf{A}_{-j}(t')\mathbf{D}_j^{\dagger}(t')\right)^{-1} \\ \mathbf{W}_3(t') = \mathbf{D}_j(t')\mathbf{A}_{-j}(t')\mathbf{A}_j(t')^{-1} \end{cases} \tag{25}$$

$t' \geq 0$ denotes the t'-th iteration, and $\mathbf{A}_j(t') := \sum_{i \in \mathscr{J}} \mathbf{H}_{j,i} \mathbf{Q}_i(t') \mathbf{H}_{j,i}^{\dagger}$, $\mathbf{A}_{-j}(t') := \sum_{i \in \mathscr{J}_{-j}} \mathbf{H}_{j,i} \mathbf{Q}_i(t') \mathbf{H}_{j,i}^{\dagger}$. According to the iterative equation (24), a new iterator $\mathbf{D}_j(t'+1)$ can be yielded once the covariance matrix of transmission power at the previous iteration t', $\mathbf{Q}_j(t')$, is provided, and the previous $\mathbf{D}_j(t')$ is recorded. Thus, a series of iterators, $\{\mathbf{D}_j(t') | t' \geq 0\}$, can be achieved by an iterative algorithm, which will converge to an optimum $\widehat{\mathbf{Q}}_j$.

B. Linear Semi-definite Programming Model

Once a new iterator of $\mathbf{D}_j$ is obtained from (24), we can keep this decision variable at this iteration, i.e., treating it as a given or fixed parameter, so as to maximize the energy efficiency by a local optimum of $\mathbf{Q}_j$ obtained by solving the model (18a)∼(18b). Clearly, this model is essentially different from the conventional fractional programming since the objective function, $g_j\left(\mathbf{Q}_j, \mathbf{D}_j \mid \mathscr{H}_j, \mathscr{Q}_{-j}\right)$ in (18a), includes the collection of metrics of other secondary links' energy efficiency. Hence, the conventional fractional programming algorithm based on Dinkelbach's method in [21], [33] would fail to deal with the model. In addition, it can be seen that

(14d) is a nonlinear inequality constraint, which increases the complexity of of this model. Therefore, we do not directly solve the model with respect to $\mathbf{Q}_j$. Instead, we approach to the optimum of the original model through constructing as well as solving a sequence of first-order semi-definite programming sub-problems. The approach proposed in this work is similar to some extent to the well-known sequential linear programming (SLP) that is usually adopted for solving large-scale nonlinear optimizations [34], since the first-order linearized approximations of constraints are adopted in both methods. But, the conventional SLP uses a nonlinear objective function in its successive subproblems, so that these sub-problems are not fully linear programming, whose solution may require the support of some other well-specified complex nonlinear algorithms. Differently, in our work, the objective function as well as the involved nonlinear constraints are linearized by the first-order Taylor's approximation.

Let h_i denote $h_i = r_i - r_i^{\min}$, such that the constraint (14d) is equivalent to $h_i \geq 0$ for $\forall i \in \mathscr{J}$. In order to distinguish the iteration index t', we introduce a local time indicator, $t \geq 0$, which is used to denote the t-th iteration when an individual secondary user locally approaches the optimal $\mathbf{Q}_j$. When a new iterator, $\mathbf{D}_j(t'+1)$, is derived from (24) at the end of the t' iteration, we fix $\mathbf{D}_j(t) = \mathbf{D}_j(t'+1)$ for $\forall t \geq 0$ and initialize the decision variable $\mathbf{Q}_j(t)$ as $\mathbf{Q}_j(0) = \mathbf{Q}_j(t')$. Accordingly, we can further approximate this condition by the first-order Taylor expansion. We assume that $(\mathbf{Q}_j(t), \mathbf{D}_j(t))$ is the feasible solution for the model in (18a)$\sim$(18b) obtained at the t-th iteration. The corresponding linear approximation of h_i with respect to $\mathbf{Q}_j$ at the iterator $(\mathbf{Q}_j(t), \mathbf{D}_j(t))$ is

$$\begin{aligned}\widetilde{h}_i\left(\mathbf{Q}_j\right) \approx & \left(r_i\left(\mathbf{Q}_j(t), \mathbf{D}_j(t)\right) - r_i^{\min}\right) \\ & + \operatorname{vec}^T\left(\frac{\partial r_i(\mathbf{Q}_j(t), \mathbf{D}_j(t))}{\partial \mathbf{Q}_j^*}\right) \Delta \mathbf{Q}_j(t)\end{aligned} \tag{26}$$

for $\forall i \mathscr{J}$ where $r_i\left(\mathbf{Q}_j(t), \mathbf{D}_j(t)\right)$ denotes the transmission rate r_i evaluated at the given point $(\mathbf{Q}_j(t), \mathbf{D}_j(t))$, and $\Delta \mathbf{Q}_j(t)$ denotes $\Delta \mathbf{Q}_j(t) = \operatorname{vec}(\mathbf{Q}_j) - \operatorname{vec}(\mathbf{Q}_j(t))$. In particular, we have the following different expressions for the complex gradient of r_i with respect to $\mathbf{Q}_j$

$$\frac{\partial r_i}{\partial \mathbf{Q}_j^*} = \begin{cases} \frac{1}{\ln 2}\left\{\mathbf{H}_{i,j}^{\dagger}\mathbf{D}_i^{\dagger}\mathbf{\Gamma}(i,j)\mathbf{D}_i\mathbf{H}_{i,j}\right\}, i \neq j & \text{(27a)} \\ \frac{1}{\ln 2}\left\{\mathbf{H}_{j,j}^{\dagger}\mathbf{D}_j^{\dagger}\mathbf{Z}(j)^{-1}\mathbf{D}_j\mathbf{H}_{j,j}\right\}, i = j & \text{(27b)} \end{cases}$$

where we define

$$\begin{aligned}\mathbf{\Gamma}(i,j) = & \left(\mathbf{B}_{i,-j} + \mathbf{D}_i\mathbf{H}_{i,j}\mathbf{Q}_j\mathbf{H}_{i,j}^{\dagger}\mathbf{D}_i^{\dagger}\right)^{-1} \\ & - \left(\mathbf{B}_{-i,-j} + \mathbf{D}_i\mathbf{H}_{i,j}\mathbf{Q}_j\mathbf{H}_{i,j}^{\dagger}\mathbf{D}_i^{\dagger}\right)^{-1}\end{aligned} \tag{28}$$

$$\mathbf{Z}(j) = \mathbf{\Phi}\left(\mathbf{D}_j, \mathscr{Q}_{-j}\right) + \mathbf{D}_j\mathbf{H}_{j,j}\mathbf{Q}_j\mathbf{H}_{j,j}^{\dagger}\mathbf{D}_j^{\dagger}. \tag{29}$$

$\mathbf{B}_{i,-j}$ and $\mathbf{B}_{-i,-j}$ are defined as follows, respectively,

$$\mathbf{B}_{i,-j} := \widetilde{\mathbf{R}}_i + \sum_{k \in \mathscr{J}_{-j}} \mathbf{D}_i\mathbf{H}_{i,k}\mathbf{Q}_k\mathbf{H}_{i,k}^{\dagger}\mathbf{D}_i^{\dagger} \tag{30}$$

$$\mathbf{B}_{-i,-j} := \widetilde{\mathbf{R}}_i + \sum_{k \in \mathscr{J}_{-i} \cap \mathscr{J}_{-j}} \mathbf{D}_i\mathbf{H}_{i,k}\mathbf{Q}_k\mathbf{H}_{i,k}^{\dagger}\mathbf{D}_i^{\dagger} \tag{31}$$

Noting that $\frac{\partial \mathrm{Tr}(\mathbf{X})}{\partial \mathbf{X}^*} = \mathbf{0}$ is held for any complex square matrix $\mathbf{X} \in \mathbb{C}$, we find $\frac{\partial p_j}{\partial \mathbf{Q}_j^*} = \mathbf{0}_{m_j \times m_j}$. Thus, using the chain rule we can further yield $\bigtriangledown_{\mathbf{Q}_j} f_j\left(\mathbf{Q}_j, \mathbf{D}_j \mid \mathscr{H}_j, \mathscr{Q}_{-j}\right) = \frac{1}{p_j}\bigtriangledown_{\mathbf{Q}_j} r_j$, which leads to

$$\bigtriangledown_{\mathbf{Q}_j} g_j(\mathbf{Q}_j, \mathbf{D}_j \mid \mathscr{H}_j, \mathscr{Q}_{-j}) = \sum_{i \in \mathscr{J}} \frac{1}{p_i}\frac{\partial r_i}{\partial \mathbf{Q}_j^*}. \tag{32}$$

Following (27a), (27b) and (32) above we can approximate the objective function (18a) with the first-order Taylor's expansion at the iterator point $(\mathbf{Q}_j(t), \mathbf{D}_j(t))$, which can be expressed as following

$$\begin{aligned}\widetilde{g}_j(\mathbf{Q}_j, \mathbf{D}_j(t) \mid \mathscr{H}_j, \mathscr{Q}_{-j}(t)) \approx & g_j(\mathbf{Q}_j(t), \mathbf{D}_j(t) \mid \mathscr{H}_j, \mathscr{Q}_{-j}(t)) \\ & + \mathbf{G}_j\left(\mathbf{Q}_j(t), \mathbf{D}_j(t)\right) \Delta \mathbf{Q}_j(t)\end{aligned} \tag{33}$$

where $\mathbf{G}_j\left(\mathbf{Q}_j(t), \mathbf{D}_j(t)\right)$ is used to denote the transposed vectorized complex-valued gradient of $g_j(\mathbf{Q}_j, \mathbf{D}_j \mid \mathscr{H}_j, \mathscr{Q}_{-j})$ evaluated at the point $(\mathbf{Q}_j(t), \mathbf{D}_j(t))$. That is, we represent $\mathbf{G}_j\left(\mathbf{Q}_j(t), \mathbf{D}_j(t)\right)$ as

$$\mathbf{G}_j\left(\mathbf{Q}_j(t), \mathbf{D}_j(t)\right) = \operatorname{vec}^T\left(\bigtriangledown_{\mathbf{Q}_j} g_j(\mathbf{Q}_j(t), \mathbf{D}_j(t) \mid \mathscr{H}_j, \mathscr{Q}_{-j})\right). \tag{34}$$

According to the linear approximations of the objective function, $g_j(\mathbf{Q}_j, \mathbf{D}_j \mid \mathscr{H}_j, \mathscr{Q}_{-j})$, and the difference between the transmission rate r_j and the minimum $r_i^{\min}$, h_i, we can arrive at a linear programming subproblem for the j-th secondary link at the t-th iteration as follows

$$\underset{\{(\mathbf{Q}_j, \lambda_j)\}}{\text{maximize}} : \widetilde{g}_j\left(\mathbf{Q}_j, \mathbf{D}_j(t) \mid \mathscr{H}_j, \mathscr{Q}_{-j}(t)\right) \tag{35a}$$

$$s.t.\ \mathbf{Q}_j \in \mathscr{R}_j\left(\mathbf{Q}_j, \lambda_j\right) \tag{35b}$$

where $\mathscr{R}_j\left(\mathbf{Q}_j, \lambda_j\right)$ denotes the linearized constraint set corresponding to $\mathbf{Q}_j$, defined by

$$\mathscr{R}_j\left(\mathbf{Q}_j, \lambda_j\right) := \left\{ (\mathbf{Q}_j, \lambda_j) \left| \begin{array}{l} (14b) \\ (17) \\ \widetilde{h}_i(\mathbf{Q}_j) \geq 0 \text{ for } \forall i \in \mathscr{J} \\ \mathbf{Q}_j \in \mathbb{M}_+^{m_j \times m_j} \end{array} \right. \right\} \tag{36}$$

This sub-optimization problem (35) is a type of linear approximation of the original model (18). In (35), the main decision variables are $\mathbf{Q}_j$ and λ_j, while the other secondary users' complex-valued coding matrices are fixed at the t-th iteration, i.e., $\mathbf{Q}_i \in \mathscr{Q}_{-j}(t) := \{\mathbf{Q}_i(t), i \in \mathscr{J}_{-j}\}$. The complex-valued decoding matrices are also kept at the same iteration, i.e., $\{\mathbf{D}_i = \mathbf{D}_i(t), i \in \mathscr{J}\}$. Furthermore, we have to remark that since the first-order Taylor's expansion is adopted to approximate the nonlinear constraint condition, $\widetilde{h}_j(\mathbf{Q}_j)$, as shown in (26), an optimum point derived by solving the linear programming problem (35) above may be an infeasible solution for the original model (18). Indeed, there could exist a slight gap between the first-order approximation at the point $(\mathbf{Q}_j(t), \mathbf{D}_j(t))$, $\widetilde{h}_i(\mathbf{Q}_j)$, and its original form h_i, such that a point, $\mathbf{Q}_j$, even though guaranteeing $\widetilde{h}_i(\mathbf{Q}_j) \geq 0$, may not satisfy $h_i \geq 0$. Therefore, we do not directly solve the model (35). Instead, to obtain a descent feasible iteration direction, we first rewrite the maximum power consumption constraint

(14b) using the transformation formula for connecting the $\mathrm{Tr}(\cdot)$ and $\mathrm{vec}(\cdot)$ operators, $\mathrm{Tr}(\mathbf{A}^T\mathbf{B}) = \mathrm{vec}^T(\mathbf{A})\mathrm{vec}(\mathbf{B})$ (where $\mathbf{A} \in \mathbb{C}^{n\times m}$ and $\mathbf{B} \in \mathbb{C}^{n\times m}$), as

$$p_j^{\max} - p_{j,c} - \beta \mathrm{vec}^T(\mathbf{I}_{m_j})\mathrm{vec}(\mathbf{Q}_j) \geq 0 \tag{37}$$

where $\mathbf{I}_{m_j}$ denotes an $m_j \times m_j$ identity matrix. Hence, we represent the left term of the inequality (37) by $w_j(\mathbf{Q}_j) := p_j^{\max} - p_{j,c} - \beta \mathrm{vec}^T(\mathbf{I}_{m_j})\mathrm{vec}(\mathbf{Q}_j)$, and let $c_j := p_j^{\max} - p_{j,c} - \beta \mathrm{vec}^T(\mathbf{I}_{m_j})\mathrm{vec}(\mathbf{Q}_j(t))$ and $\mathbf{d}_j(t) := \mathbf{Q}_j - \mathbf{Q}_j(t)$. We can further re-express $w_j(\mathbf{Q}_j)$ as the function of $\mathbf{d}_j(t)$, i.e., $w_j(\mathbf{d}_j(t)) = w_j(\mathbf{Q}_j) = c_j - \beta \mathrm{vec}^T(\mathbf{I}_{m_j})\mathrm{vec}(\mathbf{d}_j(t))$. Similarly, the first-order approximation, $\widetilde{h}_i(\mathbf{Q}_j)$, can be easily rewritten as $\widetilde{h}_i(\mathbf{d}_j(t)) = \widetilde{h}_i(\mathbf{Q}_j) = h_i(\mathbf{Q}_j(t)) + \bigtriangledown_{\mathbf{Q}_j} h_i(\mathbf{Q}_j(t))\mathrm{vec}(\mathbf{d}_j(t))$ for $\forall i \in \mathscr{J}$ where $h_i(\mathbf{Q}_j(t)) = r_i(\mathbf{Q}_j(t), \mathbf{D}_j(t)) - r_i^{\min}$ and

$$\bigtriangledown_{\mathbf{Q}_j} h_i(\mathbf{Q}_j(t)) = \mathrm{vec}^T\left(\frac{\partial r_i(\mathbf{Q}_j(t), \mathbf{D}_j(t))}{\partial \mathbf{Q}_j^*}\right). \tag{38}$$

To simplify the expression we also introduce the matrix notation, $LM_j(\mathbf{d}_j(t), \lambda_j)$

$$LM_j(\mathbf{d}_j(t), \lambda_j) = \begin{bmatrix} LM_{j,11} & LM_{j,12} \\ LM_{j,21} & LM_{j,22} \end{bmatrix} \tag{39}$$

whose components are defined by

$$\begin{cases} LM_{j,11} = \lambda_j \mathbf{I}_{m_j u} - (\mathbf{I}_u \otimes \mathbf{Q}_j(t)) - (\mathbf{I}_u \otimes \mathbf{d}_j(t)) \\ LM_{j,12} = -\mathrm{vec}\left(\mathbf{d}_j(t)^\dagger \widetilde{\mathbf{U}}_j^\dagger\right) - \mathrm{vec}\left(\mathbf{Q}_j(t)^\dagger \widetilde{\mathbf{U}}_j^\dagger\right) \\ LM_{j,21} = -\left(\mathrm{vec}\left(\mathbf{d}_j(t)^\dagger \widetilde{\mathbf{U}}_j^\dagger\right) + \mathrm{vec}\left(\mathbf{Q}_j(t)^\dagger \widetilde{\mathbf{U}}_j^\dagger\right)\right)^\dagger \\ LM_{j,22} = u_j(t) - \mathrm{vec}^T(\mathbf{d}_j(t))\left[\mathbf{I}_{m_j} \otimes \widetilde{\mathbf{U}}_j^T\right]\mathrm{vec}\left(\widetilde{\mathbf{U}}_j^*\right) \end{cases} \tag{40}$$

and $u_j(t) := \gamma_j^{\max} - \lambda_j \sigma_j^2 - \mathrm{Tr}\left(\widetilde{\mathbf{U}}_j \mathbf{Q}_j(t) \widetilde{\mathbf{U}}_j^\dagger\right)$. We remark that the linear matrix $LM_j(\mathbf{d}_j(t), \lambda_j)$ is indeed equivalent to the left term of (17).

Since the original problem is formulated as a maximum optimization model as given in (18), a series of iterators should be generated in such a direction that makes the objective function increasing. At this point, we just call this direction an 'ascent direction'. According to the form of the linear programming model (36), we can refer to the method of feasible directions of Topkis-Veinott to obtain an ascent feasible direction at the t-th iteration. Using the notations introduced above, $\mathbf{G}_j(\mathbf{Q}_j(t), \mathbf{D}_j(t))$, $\widetilde{h}_i(\mathbf{d}_j(t))$, $w_j(\mathbf{d}_j(t))$ and $LM_j(\mathbf{d}_j(t), \lambda_j)$, we derive another linear programming subproblem associated with (35) as

$$\underset{\{(\mathbf{d}_j(t), \lambda_j, \omega)\}}{\text{minimize}} : \omega \tag{41a}$$

$$s.t.\ \mathbf{G}_j(\mathbf{Q}_j(t), \mathbf{D}_j(t))\,\mathrm{vec}(\mathbf{d}_j(t)) + \omega \geq 0 \tag{41b}$$

$$\widetilde{h}_i(\mathbf{d}_j(t)) + \omega \geq 0, i \in \mathscr{J} \tag{41c}$$

$$w_j(\mathbf{d}_j(t)) \geq 0 \tag{41d}$$

$$LM_j(\mathbf{d}_j(t), \lambda_j) \succeq \mathbf{0}_{(m_j u+1)} \tag{41e}$$

$$\left|[\mathbf{d}_j(t)]_{l_1,l_2}\right| \leq 1, l_1, l_2 = 1, 2, \ldots, m_j \tag{41f}$$

$$\lambda_j \geq 0 \tag{41g}$$

where ω is a auxiliary variable and $[\mathbf{d}_j(t)]_{l_1,l_2}$ represents the (l_1, l_2)-th entity in the matrix $\mathbf{d}_j(t)$. The linear constraint (41f) added in the model above can ensure that the searching space with respect to the decision variable $\mathbf{d}_j(t)$ is limited, such that this model can guarantee the existence of a finite optimal solution.

C. Calculation of Optimal Searching Step-size

Based on the model (41), we further formulate a one-dimension searching problem to determine the optimal step size in the direction $\mathbf{d}_j(t)$ obtained by solving (41) at the t-th iteration. We denote the step size by $\Lambda_j(t) \in \mathbb{R}_+$. With this notation, we propose the following model to determine the optimal $\Lambda_j(t)$:

$$\underset{\{\Lambda_j(t)\}}{\text{maximize}} : g_j(\mathbf{Q}_j(t) + \Lambda_j(t)\mathbf{d}_j(t), \mathbf{D}_j(t) \mid \mathscr{H}_j, \mathscr{Q}_{-j}(t)) \tag{42a}$$

$$s.t.\ 0 \leq \Lambda_j(t) \leq \Lambda_j^{\max}(t) \tag{42b}$$

where the upper bound of $\Lambda_j(t)$ can be constructed as

$$\Lambda_j^{\max}(t) := sup\left\{\Lambda_j(t) \;\middle|\; \begin{array}{l} w_j(\mathbf{Q}_j(t) + \Lambda_j(t)\mathbf{d}_j(t)) \geq 0 \\ h_i(\mathbf{Q}_j(t) + \Lambda_j(t)\mathbf{d}_j(t)) \geq 0, \forall i \\ LM_j(\Lambda_j(t)\mathbf{d}_j(t), \lambda_j) \succeq \mathbf{0} \end{array}\right\} \tag{43}$$

Once the optimal $\Lambda_j(t)$ is derived by solving the sub-problem (42) above, a new iterator approaching the optimal $\mathbf{Q}_j$ at the next step can be constructed by the increment equation

$$\mathbf{Q}_j(t+1) = \mathbf{Q}_j(t) + \Lambda_j(t)\mathbf{d}_j(t) \tag{44}$$

We have to point out that the models (41) and (42) are locally solved by the individual secondary user of the cognitive radio network, which can be implemented in an on-linear manner. Recalling the definitions of the iteration indicators t' and t, we can set a new iterator of the decision variable $\mathbf{Q}_j$ at the end of the $(t'+1)$-th time slot as

$$\mathbf{Q}_j(t'+1) = \mathbf{Q}_j(t^*) \tag{45}$$

where t^* is used to indicate the iteration at which $\{\mathbf{Q}_j(t), t \geq 0\}$ converges. Let $\epsilon > 0$ be a sufficiently small real number that is a pre-defined threshold. Then, we present a simple convergence condition for terminating the local iterations as $|\omega(t^*)| \leq \epsilon$ where $\omega(t^*)$ denotes the optimal value of the objective function of the model (41) constructed at the t^*-th iteration.

D. Robust Distributed Algorithm

Generally, an energy-efficient transmission optimization scheme implemented in a distributed manner is appealing for the realization of a high-scalable robust cognitive radio network. Recalling the formulas given in (24) and (45), $\mathbf{Q}_j(t')$ and $\mathbf{D}_j(t')$ can be updated in turn at each t'-th iteration in a on-line way, and the proposed iterative approach can induce a distributed optimization procedure. Specifically, as shown in (24), each secondary transmitter of the CR network Tx_j ($j \in \mathscr{J}$) can perform the iterative procedure for updating $\mathbf{D}_j$ locally based on its own noise covariance matrix $\widetilde{\mathbf{R}}_j$ and

interference channel matrix $\mathbf{H}_{j,j}$, the recorded iterator $\mathbf{D}_j(t')$ and the covariance matrix $\mathbf{Q}_j(t')$ derived at the previous iteration t', and this node can also exploit the other information of its neighboring links, including the links' covariance matrices recorded at t', $\mathscr{Q}_{-j}(t')$, and the interference channel matrices, $\mathscr{H}_j$, to calculate the parameters $\mathbf{A}_j(t')$ and $\mathbf{A}_{-j}(t')$.

We remark that the local message passing is needed to realize the distributed algorithm. The information can be obtained from message broadcasting via feedback channel. Moreover, it should be pointed out that when some secondary links whose locations are far away from the j-th users', the existing path loss effect can lead to quite small channel interference arising from such remote links. Thus, the product terms corresponding to those links involved in $\mathbf{A}_j(t')$ and $\mathbf{A}_{-j}(t')$ can be properly neglected. This implies that the information needed to computing $\mathbf{A}_j(t')$ and $\mathbf{A}_{-j}(t')$ can be limited to the local neighbors of the j-th secondary user, such that only local $\left\{\mathbf{H}_{j,i}, i \in \mathscr{J}_{-j}^{local}\right\}$ and $\left\{\mathbf{Q}_i(t'), i \in \mathscr{J}_{-j}^{local}\right\}$ should be exchanged ($\mathscr{J}_{-j}^{local}$ is used to denote the set of neighboring links associated with the j-th secondary user). Once a new decoding matrix $\mathbf{D}_j(t'+1)$ is obtained by Tx_j, a new iterator of $\mathbf{Q}_j$, $\mathbf{Q}_j(t'+1)$, can also be yielded by sequentially solving the linear programming sub-models (41) and (42) based on the new iterator of $\mathbf{D}_j$, $\mathbf{D}_j(t'+1)$ and neighboring secondary links' covariance matrix and interference channel matrix information at t'. Then, the next iterator can also be derived based on the pair $(\mathbf{Q}_j(t'+1), \mathbf{D}_j(t'+1))$. The iterative procedure should be repeated until a certain prespecified stopping criterion is satisfied, such as

$$\left| \begin{array}{l} g_j\left(\mathbf{Q}_j(\tilde{t}'), \mathbf{D}_j(\tilde{t}') \mid \mathscr{H}_j, \mathscr{Q}_{-j}(\tilde{t}')\right) \\ -g_j\left(\mathbf{Q}_j(\tilde{t}'-1), \mathbf{D}_j(\tilde{t}'-1) \mid \mathscr{H}_j, \mathscr{Q}_{-j}(\tilde{t}'-1)\right) \end{array} \right| \leq \varepsilon \tag{46}$$

where $\tilde{t}' \geq t'$ denotes a certain time instant when a series of the pairs $\{(\mathbf{Q}_j(t'), \mathbf{D}_j(t'))\}$ converge, and ε is a well predefined positive real number that should be sufficiently small. The distributed algorithm is summarized in **Algorithm 1**.

It should be noted that the cross-channel matrix information may be dynamically changing in an actual implementation scenario. In **Algorithm 1**, slow changes of the cross-channel matrices can be taken into consideration. That is, when any change in the cross-channel matrices is sensed, new cross-channel matrices should be recorded and re-broadcasted to the corresponding users.

Algorithm 1 Robust distributed EE transmission optimization

Initialize $(\mathbf{Q}_i(0), \mathbf{D}_i(0))$ for $\forall i$ and let $t' \leftarrow 0$
repeat for $t' = 0, 1, 2, \ldots$
  for all $j \in \mathscr{J}$ **do**
    Tx_j broadcasts $(\mathbf{Q}_j(t'), \mathbf{D}_j(t'))$
    Tx_j collects $\left\{(\mathbf{Q}_i(t'), \mathbf{D}_i(t')), i \in \mathscr{J}_{-j}^{local}\right\}$
    Tx_j re-broadcasts $\left\{(\mathbf{Q}_i(t'), \mathbf{D}_i(t')), i \in \mathscr{J}_{-j}^{local}\right\}$
    Tx_j obtains $\left\{\mathbf{H}_{j,i}, i \in \mathscr{J}_{-j}^{local} \cup \{j\}\right\}$
    Compute $\mathbf{A}_j(t')$ and $\mathbf{A}_{-j}(t')$
    Compute $\mathbf{D}_j(t'+1)$ based on (24)
    Update $\mathbf{D}_j(t')$ by $\mathbf{D}_j(t') \leftarrow \mathbf{D}_j(t'+1)$
    Locally initialize $t \leftarrow 0$, $\mathbf{Q}_j(0) \leftarrow \mathbf{Q}_j(t')$
    Keep $\mathbf{D}_j(t) = \mathbf{D}_j(t'+1)$ for $\forall t \leq 0$
    repeat for $t = 0, 1, 2, \ldots$
      Solve an optimal $(\mathbf{d}_j(t), \lambda_j, \omega)$ from (41)
      Solve an optimal $\Lambda_j(t)$ from (42)
      Update $\mathbf{Q}_j(t)$ by $\mathbf{Q}_j(t) \leftarrow \mathbf{Q}_j(t) + \Lambda_j(t)\mathbf{d}_j(t)$
    until $|\omega| \leq \epsilon$ for Tx_j
    Tx_j sets $\mathbf{Q}_j(t'+1) \leftarrow \mathbf{Q}_j(t)$
    Tx_j updates $\mathbf{Q}_j(t')$ by $\mathbf{Q}_j(t') \leftarrow \mathbf{Q}_j(t'+1)$
    Tx_j feedbacks $(\mathbf{Q}_j(t'), \mathbf{D}_j(t'))$ to Rx_j
  end for
until Stopping criterion (46) is satisfied for $j \in \mathscr{J}$

IV. Theoretical Analysis

The feasibility and the convergence of a optimization algorithm is of paramount significance in terms of theoretical analysis on its performance. Thus, we first establish a lemma associated the proposed model (See the Lemma 1 in Appendix A-A). Based on the Lemma 1, we can show the feasibility of a series of iteration directions obtained by solving the subproblem (41):

Theorem 1. *Suppose that $\mathscr{T}$ is a finite set of discrete iterator indexes. Any subsequence of $\{\mathbf{d}_j(t), t \in \mathscr{T}\}$ generated by solving the corresponding linear programming sub-problem (41) are feasible improving iteration directions.*

Proof: According to the lemma 1, constructing a feasible iteration direction is equivalent to solving a group of the inequalities (56), (57), (58) and (59). Furthermore, taking into account the inactive constraints at the point $\mathbf{Q}_j(t)$,

$$\{h_i(\mathbf{Q}_j(t)) > 0, i \notin A_j(\mathbf{Q}_j(t))\} \tag{47}$$

and the active constraints at the same point,

$$\{h_i(\mathbf{Q}_j(t)) = 0, i \in A(\mathbf{Q}_j(t))\} \tag{48}$$

we can modify the condition (57) as

$$\left\{ \begin{array}{l} \left(\begin{array}{l} h_i(\mathbf{Q}_j(t)) \\ + \bigtriangledown_{\mathbf{Q}_j} h_i\left(\mathbf{Q}_j(t)\right) \mathrm{vec}(\mathbf{d}_j(t)) \end{array} \right) > 0, i \in A(\mathbf{Q}_j(t)) \quad (49) \\ h_i(\mathbf{Q}_j(t)) > 0, i \notin A(\mathbf{Q}_j(t)) \quad (50) \end{array} \right.$$

Note that the left side of the inequality (49) is equal to $\widetilde{h}_i(\mathbf{d}_j(t))$ $(i \in A(\mathbf{Q}_j(t)))$, and (50) can immediately lead to $\widetilde{h}_i(\mathbf{d}_j(t)) > 0$ $(i \notin A(\mathbf{Q}_j(t)))$. We can re-express the conditions (49) and (50) by a unified form

$$\widetilde{h}_i(\mathbf{d}_j(t)) > 0, \forall i \in \mathscr{J} \tag{51}$$

Substituting the condition (57) with (51), we can find that solving a group of linear inequalities (56), (51), (58) and (59) is equivalent to solving the corresponding linear programming problem as given in (41). Hence, we prove this theorem. ■

Next, in order to establish the convergence of any sequence of iterators generated by our proposed algorithm to a Fritz John point, we provide a corollary that outlines basic properties of any iterator sequence. For simplicity, let $\mathbf{\Xi}$ denote the feasible solution space of the covariance matrix $\mathbf{Q}_j$, which is a non-empty closed set in $\mathbb{M}_+^{m_j \times m_j}$, i.e., $\mathbf{\Xi} \subset \mathbb{M}_+^{m_j \times m_j}$. Then, without loss of generality, let $\mathbf{Q}_j \in \mathbf{\Xi}$ be a feasible decision variable of the original model (18) and $\mathbf{d}_j$ an improving

feasible direction of the objective function g_j at $\mathbf{Q}_j$. Supposing that Λ_j is an optimal solution of the one-dimension searching problem (42) where $\mathbf{Q}_j + \Lambda_j \mathbf{d}_j \in \mathbf{\Xi}$, we get a corollary associated with the model (41) as follows:

Corollary 2. *Consider that $\{\mathbf{d}_j(t)\}$ is any sequence of iteration directions derived by the model (41) and $\{\mathbf{Q}_j(t)\}$ the corresponding sequence of decision points. None of potential subsequences $\{(\mathbf{Q}_j(t), \mathbf{d}_j(t)), t \in \mathscr{T}\}$ generated from (41) and (44) can satisfy all the following four conditions where $\mathscr{T}$ denotes a finite discrete time set: (i) $\mathbf{d}_j(t) \to \mathbf{d}_j$ for $t \in \mathscr{T}$; (ii) $\mathbf{Q}_j(t) \to \mathbf{Q}_j$ for $t \in \mathscr{T}$; (iii) $\mathbf{Q}_j(t) + \Lambda_j \mathbf{d}_j(t) \in \mathbf{\Xi}$ for $\forall \Lambda_j \in [0, \Lambda_j']$ and for $t \in \mathscr{T}$ where Λ_j' represents a certain bound of Λ_j and $\Lambda_j' > 0$; (iv) $\mathbf{G}_j(\mathbf{Q}_j, \mathbf{D}_j(t)) \operatorname{vec}(\mathbf{d}_j) > 0$ where $\mathbf{G}_j(\mathbf{Q}_j, \mathbf{D}_j(t))$ is given in (34).*

Proof: See Appendix A-D. ∎

Based on the model (41), the sufficient condition of an iterator $\mathbf{Q}_j(t)$ that is a Fritz John point can be established, which is given in the Lemma 2 (See Appendix A-B). Combining the result in the Corollary 2 and the Lemma 2, we are ready to establish the convergence of the proposed approach to a Fritz John point in the Theorem 2.

Theorem 2. *Consider that $\{\mathbf{Q}_j(t)\}$ is any sequence of iterators derived by sequentially solving the model (41) and $\{\mathbf{Q}_j(t)\}$. A limit point of $\{\mathbf{Q}_j(t)\}$ is a Fritz John point.*

Proof: Let $\mathscr{T}$ be a finite discrete time set and $\{\mathbf{Q}_j(t), t \in \mathscr{T}\}$ a subsequence of $\{\mathbf{Q}_j(t)\}$ converging to a limit point $\mathbf{Q}_j$, i.e., $\mathbf{Q}_j(t) \to \mathbf{Q}_j$ for $t \in \mathscr{T}$. By contradiction we assume that $\mathbf{Q}_j$ does not belong to a Fritz John point. Then, according to the lemma 2, the corresponding optimal value of the objective function in the model (41) obtained at $\mathbf{Q}_j$, denoted by ω, is not equal to zero. At this point, there exists a small positive real number $\tau > 0$ such that $\omega = -\tau$. Let $(\omega(t), \mathbf{d}_j(t))$ be an optimal solution of the model (41). Since $\mathscr{T}$ is finite, any $\{\mathbf{d}_j(t), t \in \mathscr{T}\}$ is bounded. Thus, there will be at least a subsequence, $\{\mathbf{d}_j(t), t \in \mathscr{T}^*\}$ where $\mathscr{T}^* \subset \mathscr{T}$, that have a limit point $\mathbf{d}_j$, i.e., $\mathbf{d}_j(t) \to \mathbf{d}_j$ for $t \in \mathscr{T}^*$. In addition, according to the continuous differentiability of g_j and h_i for all $i \in \mathscr{J}$, and $\mathbf{Q}_j(t) \to \mathbf{Q}_j$ for $t \in \mathscr{T}^*$, we can also see that for $t \in \mathscr{T}^*$ $\omega(t) \to \omega$. Thus, when $t \in \mathscr{T}^*$ is sufficiently large, we can get $\omega(t) < \frac{\omega}{N} = \frac{-\tau}{N} < 0$ where assuming $N > 1$. According to the model (41), we also have

$$\begin{cases} \mathbf{G}_j(\mathbf{Q}_j(t), \mathbf{D}_j(t)) \operatorname{vec}(\mathbf{d}_j(t)) \geq -\omega(t) > \dfrac{\tau}{N} & (52) \\ \widetilde{h}_i(\mathbf{Q}_j(t)) = h_i(\mathbf{Q}_j(t)) + \bigtriangledown_{\mathbf{Q}_j} h_i(\mathbf{Q}_j(t)) \operatorname{vec}(\mathbf{d}_j(t)) \\ \qquad \geq -\omega(t) > \dfrac{\tau}{N}, \forall i \in \mathscr{J} & (53) \end{cases}$$

for $t \in \mathscr{T}^*$ sufficiently large. Since the result of (52), it follows $\mathbf{G}_j(\mathbf{Q}_j, \mathbf{D}_j(t)) \operatorname{vec}(\mathbf{d}_j) > 0$.

Let M be a real number and larger than 1, i.e., $M > 1$. The continuous differentiability of h_i and the result of (53) imply that there must be a positive real number denoted by $\Lambda_j' > 0$ such that

$$h_i(\mathbf{Q}_j(t)) + \bigtriangledown_{\mathbf{Q}_j} h_i(\mathbf{Q}_j(t) + \Lambda_j \mathbf{d}_j(t)) \operatorname{vec}(\mathbf{d}_j(t)) > \frac{\tau}{NM} \quad (54)$$

is held for any $\Lambda_j \in [0, \Lambda_j']$, for $t \in \mathscr{T}^*$ sufficiently large and for all $i \in \mathscr{T}^*$.

Subsequently, let Λ_j be bounded on $[0, \Lambda_j']$. Applying the mean value theorem can arrive at $h_i(\mathbf{Q}_j(t) + \Lambda_j \mathbf{d}_j(t)) = h_i(\mathbf{Q}_j(t)) + \Lambda_j \bigtriangledown_{\mathbf{Q}_j} h_i(\mathbf{Q}_j(t) + \xi \Lambda_j \mathbf{d}_j(t)) \operatorname{vec}(\mathbf{d}_j(t))$ for $t \in \mathscr{T}^*$ and all $i \in \mathscr{J}$ where $\xi \in (0, 1)$. Recalling $h_i(\mathbf{Q}_j(t)) \geq 0$ for $t \in \mathscr{T}^*$ and all $i \in \mathscr{J}$ where $\xi \in (0, 1)$, $\xi \Lambda_j \in [0, \Lambda_j']$, and since (53) and (54), we can get

$$h_i(\mathbf{Q}_j(t) + \Lambda_j \mathbf{d}_j(t)) \geq (1 - \Lambda_j) h_i(\mathbf{Q}_j(t)) + \Lambda_j \frac{\tau}{NM} \geq 0 \quad (55)$$

This inequality (55) shows that $\mathbf{Q}_j(t) + \Lambda_j \mathbf{d}_j(t)$ for any $\Lambda_j \in [0, \Lambda_j']$ can also be a feasible solution when $t \in \mathscr{T}^*$ is sufficiently large.

To sum up, this subsequence $\{(\mathbf{Q}_j(t), \mathbf{d}_j(t)), t \in \mathscr{T}^*\}$ is showed to satisfy all of the four properties given in the corollary 2. But, as discussed in the corollary 2, such a subsequence does not exist. Thus, the contradiction occurs, which indicates that the limit point $\mathbf{Q}_j$ should be a Fritz John point. ∎

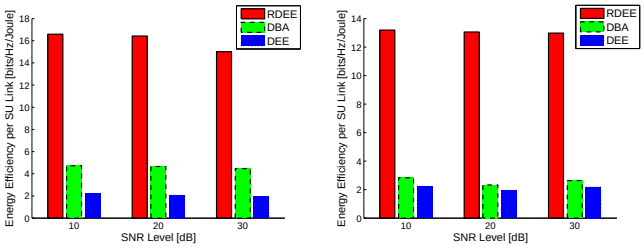
In actual numerical computation, most of the computational effort is indeed taken up in solving a sequence of linear semi-definite programming sub-problems. The MATLAB-based toolbox YALMIP [35] along with a semi-definite and second-order cone optimization solver SeDuMi [36] is used here for solving the proposed linear semi-definite programming model (41). Accordingly, SeDuMi adopts the Mehrotra-type technique of adaptive predictor-corrector algorithm with an theoretically proven $O(\sqrt{n}|\log\epsilon|)$ worst-case iteration bound (here n denotes the number of the decision variables and ϵ is a pre-specified tolerance in numerical computation), proved to be among the most efficient primal-dual interior methods in practice. At this point, it can provide a low (polynomial-time) iteration complexity for solving (41), which is expected to have an $O\left(\sqrt{m_j \times m_j + 2}|\log\epsilon|\right)$ worst-case iteration bound (noting that the model (41) includes $m_j \times m_j + 2$ decision variables).

V. Performance evaluation

In this section, we perform computer simulation experiments to evaluate the performance of the algorithm proposed in the work. For simplicity, we set the covariance matrix of the noise signal as $\widetilde{\mathbf{R}}_j = \rho_j^2 \mathbf{I}$ and $\rho_j = 0.01$ for all j where $\mathbf{I}$ is the identity matrix whose size is the same with that of $\widetilde{\mathbf{R}}_j$. We remark that there exists considerable debate in the current research work on narrowband small-scale fading statistics in V2V communication channels [37]. [37] also stated that many studies for V2V channels do not distinguish between LOS and NLOS situations. As argued in [38], it is not easy to distinguish small-scale fading from large-scale fading in V2V communications. The issue of accurately modeling realistic channels for V2V and V2I communications and some aspects relevant to channel measurement and mathematical characterization are interesting but challenging, which are really out of scope of this study. As one of the earliest references on modeling V2V channels, [39] adopted Rayleigh fading statistics so as to derive new envelope autocorrelation

functions and Doppler spectra in V2V communications. Additionally, Rayleigh fading channel has been adopted to model mobile-to-mobile communications [40], [41]. Accordingly, for the sake of example, we also adopt Rayleigh fading channels following with zero-mean complex Gaussian distribution in our simulation experiments, and assume that the path loss effect is proportional to $dist(j,i)^{-p}$ according to [26], where $dist(j,i)$ denotes the geometric distance between the transmitter Tx_j and the receiver Rx_i, and the constant p is set to 2.5. Then we assume any entry of the channel matrix $\mathbf{H}_{j,i}$ for all $j,i \in \mathscr{J}$ is independently identically distributed, such that $\mathbf{H}_{j,i}$ can be formulated as the zero-mean unit-variance circularly symmetric complex Gaussian matrix. Similarly, we also represent any $\widetilde{\mathbf{U}}_j$ as a zero-mean unit-variance circularly symmetric complex Gaussian matrix. For comparison, we consider two cases (named C.1 and C.2 respectively) in our experiments. In the first case, we consider 5 pairs of secondary links with V2V communications which co-exist with one primary user. The antenna number of those secondary users is equally set to 2, i.e., $m_j = n_j = 2$ for $\forall j \in \mathscr{J}$, while that of the primary receiver's is also 2, i.e., $u = 2$. The relative distance between any transmitter Tx_j and its desired receiver Rx_j in the group, $dist(j,j)$, is stable at 40m, while the relative distance between any secondary transmitter and the primary receiver, or between any undesired secondary pair, $dist(j,i)$ for any $j \neq i$, are uniformly randomly generated over [40,160](m). In the second case, we consider another MIMO CR-enabled vehicular network consisting of 10 pairs of secondary users all equipped with 3 antennas and one primary receiver whose antenna number is also 3. The relative distance parameter $dist(j,j)$ is kept at 50m for $\forall j \in \mathscr{J}$, while the distance of interfering links, $dist(j,i)$ for $j \neq i$, as well as the distance between any transmitter and the primary receiver is uniformly randomly distributed over [50,200](m). We set the bound parameter of uncertain channel matrix as $\sigma_j^2 = 0.5\|\widetilde{\mathbf{U}}_j\|_F^2$ for $\forall j \in \mathscr{J}$. The minimum transmission rate corresponding to each secondary link, ,$r_j^{\min}$, is set to be 2bps/Hz. The maximum interfering power generated by any secondary link, $\gamma_j^{\max}$, is set to be 0.1W. The maximum transmission power $p_j^{\max}$ is set to guarantee a certain signal-noise-ratio (SNR) level. For simplicity we assume that a given SNR level can be represented by $L_{SNR} = \frac{p_j^{\max} - p_{j,c}}{\beta \cdot \rho_j^2 \cdot dist(j,j)^p}$. Then in the experiments let the constant parameters $p_{j,c}$ and β be 1 for $\forall j \in \mathscr{J}$. Thus, we can set $p_j^{\max}$ so as to hold a certain SNR level pre-specified for each secondary links. An advanced MATLAB-based toolbox for convex and non-convex optimization [35], YALMIP, is utilized in our work to solve the sub-optimization model (41) which is indeed a typical semidefinite programming problem.

First, we vary L_{SNR} as $L_{SNR} \in \{10, 20, 30\}$(dB) in the two simulation cases and compare our proposed robust distributed energy-efficiency optimization algorithm (RDEE) with a distributed beamforming algorithm (DBA) proposed in [21] and a decentralized EE (DEE) optimization proposed in [19]. In [21], DBA is developed through transforming the original problem of the complex nonlinear non-convex energy-efficiency optimization into a series of fractional programming sub-problems, and it approaches to an optimal solution maximizing the energy-efficiency objective function via solving the fractional programming constructed at each iteration in a distributed fashion. To maximize the energy-efficiency objective function, DEE proposed in [19] essentially employs an iterative zero-gradient-based mechanism to approach the optimum of the energy-efficient transmission optimization model with consideration of MIMO interference channels. We run each algorithms with 100 replications over different L_{SNR}. In each run of an algorithm, we independently and randomly set up channel matrices following a zero-mean unit-variance circularly symmetric complex Gaussian distribution. The energy-efficiency of each secondary link at a final convergence state in two cases are obtained by averaging the results of 100 runs, and illustrated in Fig.3a and Fig.3b, respectively. Compared with the DBA and the DEE algorithms, our beamforming approach additionally takes into account the minimum transmission rate constraint in the EE optimization model, such that we can maintain the EE of each secondary link at a relatively high level even when SNR requirement increases. From Fig.3, it can be found that our proposed algorithm achieves the highest performance among the three compared algorithms in terms of the average energy efficiency under different SNR settings in both cases.

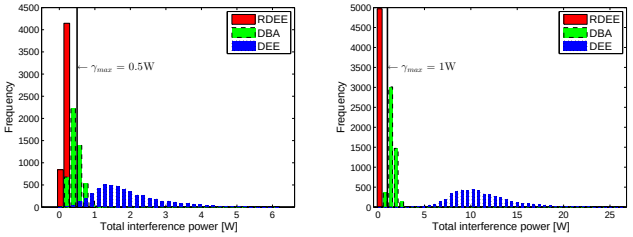


(a) C.1 (b) C.2

Fig. 3. The EE performance per secondary link for different SNR settings under the compared algorithms in the two cases.

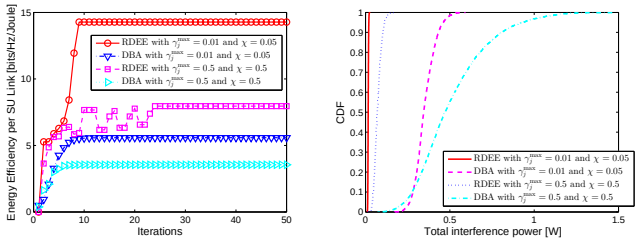
Furthermore, we keep the SNR level at 20dB and obtain different energy-efficient beamforming designs for the two cases of MIMO CR-enabled vehicular networks by applying the three algorithms. To show the robustness of our proposed solution we perform 5000 Monte Carlo simulations each of which randomly and independently realizes secondary user-to-primary channel matrices. And then we can confirm the effect of the robust interference power constraint by analyzing the frequency distribution of the total interference power received at the primary receiver. The histograms obtained in the two cases are shown in Fig.4a and Fig.4b, respectively.

Note that the tolerable interference power generated by each secondary link is set to $\gamma_j^{\max} = 0.1$W in both cases. Thus, the upper bound of the total interference power at the primary user should be $0.1 \times 5 = 0.5$W in the first cases and $0.1 \times 10 = 1$W in the second case. As illustrated in Fig.4, the interference power constraint is forced to be strictly satisfied by our proposed robust solution, while the total interference power obtained by the other two algorithms, DBA and DEE, exceeds the tolerable bound (marked by a dark vertical line in



(a) C.1 (b) C.2

Fig. 4. The frequency distribution of sum interference power under the compared algorithms in the two cases.



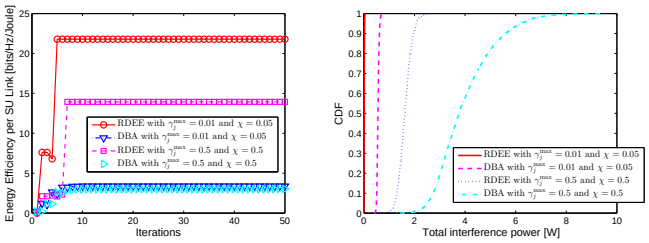
(a) The convergence of energy efficiency per secondary link. (b) The CDF of total interference power.

Fig. 5. The convergence of energy efficiency per secondary link and the CDF of total interference power obtained under different $\gamma_j^{\max}$ and χ in C.1.

these two figures) in most runs under both cases. Specifically, the figures show that the time of the DBA algorithm violating the interference bound is more than 60% in the first case and more than 80% in the second case. In particular, in most simulation instances of the DEE algorithm, the interference constraint cannot be satisfied (more than 90% of time in the first case and more than 99% in the second case). The reason is that the DEE algorithm solves the EE optimal design without consideration of the interference constraint, such that it cannot protect the primary receiver's transmission. In fact, the DBA and the DEE algorithms are not robust approach so that they frequently violates the interference power constraint. By contrast, our approach can achieve a high robustness in both experiment cases.

In order to show the influence of the uncertainty factor characterized by the parameter σ_j on the EE performance, we represent $\sigma_j^2 = \chi\|\widetilde{\mathbf{U}}_j\|_F^2$ and then vary the value of χ. In addition, we also vary the interference power bound of each secondary transmitter, $\gamma_j^{\max}$, to demonstrate its influence. Specifically, let $\chi \in \{0.05, 0.5\}$ and $\gamma_j^{\max} \in \{0.01, 0.5\}$. Now, we can compare the EE performance and convergence of our proposed algorithm with the DBA with different settings on χ and $\gamma_j^{\max}$ (We remark that since the DBA approach rather than the DEE considers the interference power constraint for EE optimization which is similar to that in our model, we mainly compare our approach with the DBA in the following experiments.). The results obtained in the first case are shown in Fig.5.

Fig.5a illustrates the convergence of EE performance metric of two algorithms. From this subfigure, we can see that although our approach converges slightly slower than the



(a) The convergence of energy efficiency per secondary link. (b) The CDF of total interference power.

Fig. 6. The convergence of energy efficiency per secondary link and the CDF of total interference power obtained under different $\gamma_j^{\max}$ and χ in C.2.

DBA, the energy efficiency per secondary link obtained by our approach with both groups of parameter settings overweighs that of the DBA. Furthermore, Fig.5b plots the cumulative distribution function (CDF) of the total interference power at the primary in order to show the robustness of both approaches. The CDF plots are also derived by running 5000 Monte Carlo simulations, in each of which we independently and randomly generate interference channel realizations. From Fig.5(5b), it can be found that the red solid and the blue dotted CDF curves of our algorithm arrives at 100% before their total interference power exceeding $5 \times 0.01 = 0.05$W and $5 \times 0.5 = 2.5$W, respectively. The magenta dashed CDF curve of the DBA shows that it seriously violates the interference power limit at 0.05W, while this algorithm can satisfy the interference constraint with a larger tolerable interference power (2.5W). In comparison, our approach has a better robustness than the DBA whenever the tolerable interference power is small or large. The similar conclusions can also be drawn from Fig.6. The convergence of the proposed RDEE and the DBA with variable $\gamma_j^{\max}$ and χ in the second experiment case is demonstrated in Fig.6a while their corresponding CDF curves obtained in the same case are given in Fig.6b. From Fig.6a, it can be found that our proposed algorithm can converge to a higher energy-efficiency level under the two groups of settings on $\gamma_j^{\max}$ and χ because our model always forces SUs to satisfy a minimum data rate. As expected, Fig.6b confirms the better robustness of our proposed approach as well. By contrast, most of the time in C.2, both the CDF curves of the non-robust DBA exceed the total interference power limits, $10\times0.01 = 0.1$W and $10\times0.5 = 5$W, respectively. In terms of both the energy-efficiency performance and the robustness, our proposed approach can provide a better beamforming design for CR-enabled vehicular MIMO transmissions.

VI. Conclusion

In this work, we have studied the robust and optimal energy-efficient beamforming design for MIMO transmissions in CR-enabled vehicular networks. We take into consideration the robust interference constraint of secondary links with imperfect CSI and the restrictions on the power budget and data rate. We have shown that even without the accurate CSI, the energy-efficiency of a secondary network can still be optimized by solving sequential linearly constrained semi-definite programming sub-problems in a robust and distributed

way. Specifically, we proposed a robust iterative optimization algorithm, the feasibility and convergence of which has been theoretically analyzed. Finally, by conducting simulations, we confirmed the strength of our method in terms of energy-efficiency performance and robustness.

APPENDIX A

A. Lemma 1

Lemma 1. *Suppose that $\mathbf{Q}_j(t)$ is one feasible solution of the original model (18) and $A(\mathbf{Q}_j(t))$ is a set of indexes of active constraints at $\mathbf{Q}_j(t)$ where $A(\mathbf{Q}_j(t)) := \{i \,|h_i\left(\mathbf{Q}_j(t)\right) = 0, i \in \mathscr{J}\}$. If $\mathbf{d}_j(t)$ satisfies the following conditions:*

$$\mathbf{G}_j\left(\mathbf{Q}_j(t), \mathbf{D}_j(t)\right) \mathrm{vec}(\mathbf{d}_j(t)) > 0 \tag{56}$$

$$\bigtriangledown_{\mathbf{Q}_j} h_i\left(\mathbf{Q}_j(t)\right) \mathrm{vec}\left(\mathbf{d}_j(t)\right) > 0, i \in A(\mathbf{Q}_j(t)) \tag{57}$$

$$w_j\left(\mathbf{d}_j(t)\right) \geq 0 \tag{58}$$

$$LM_j\left(\mathbf{d}_j(t), \lambda_j\right) \succeq \mathbf{0}_{(m_j u+1)} \tag{59}$$

then $\mathbf{d}_j(t)$ is a feasible iteration direction at the point $\mathbf{Q}_j(t)$ that can improve the value of the objective function of the original model (18).

Proof: Recalling the definitions of $w_j\left(\mathbf{d}_j(t)\right)$ and $LM_j\left(\mathbf{d}_j(t), \lambda_j\right)$, it is obvious that satisfying the conditions (58) and (59) is equivalent to guarantee the constraints (14b) and (17) in the original model (18). Thus, it reduces the proof to mainly focus on analyzing the conditions (56) and (57). Suppose the conditions (56) through (57) are held for $\mathbf{d}_j(t)$. Since $\mathbf{Q}_j(t)$ is a feasible point, it makes $h_i(\mathbf{Q}_j(t)) > 0$ held if $i \notin A(\mathbf{Q}_j(t))$. Note that $h_i(\mathbf{Q}_j(t))$ $(i \notin A(\mathbf{Q}_j(t)))$ is continuously differentiable at $\mathbf{Q}_j(t)$. Therefore, there exists a real positive number $\Lambda_j > 0$ that makes

$$h_i(\mathbf{Q}_j(t) + \Lambda_j \mathbf{d}_j(t)) \geq 0, i \notin A(\mathbf{Q}_j(t)) \tag{60}$$

when it is sufficiently small.

If $i \in A(\mathbf{Q}_j(t))$, we can derive the following equation according to the differentiability of $h_i(\mathbf{Q}_j(t))$ at $\mathbf{Q}_j(t)$

$$\begin{aligned} h_i(\mathbf{Q}_j(t) + \Lambda_j \mathbf{d}_j(t)) =& h_i(\mathbf{Q}_j(t)) \\ &+ \Lambda_j \bigtriangledown_{\mathbf{Q}_j} h_i(\mathbf{Q}_j(t)) \mathrm{vec}(\mathbf{d}_j(t)) \\ &+ \Lambda_j \|\mathrm{vec}(\mathbf{d}_j(t))\| \delta\left(\mathbf{Q}_j(t), \Lambda_j \mathbf{d}_j(t)\right) \end{aligned} \tag{61}$$

where the small term $\delta\left(\mathbf{Q}_j(t), \Lambda_j \mathbf{d}_j(t)\right)$ satisfies that when $\Lambda_j \to 0$, $\delta\left(\mathbf{Q}_j(t), \Lambda_j \mathbf{d}_j(t)\right) \to 0$, and $\|\cdot\|$ represents the Euclidean norm. Furthermore, (61) can directly lead to

$$\begin{aligned} &\frac{h_i(\mathbf{Q}_j(t) + \Lambda_j \mathbf{d}_j(t)) - h_i(\mathbf{Q}_j(t))}{\Lambda_j} \\ &= \bigtriangledown_{\mathbf{Q}_j} h_i(\mathbf{Q}_j(t)) \mathrm{vec}(\mathbf{d}_j(t)) \\ &\quad + \|\mathrm{vec}(\mathbf{d}_j(t))\| \delta\left(\mathbf{Q}_j(t), \Lambda_j \mathbf{d}_j(t)\right) \end{aligned} \tag{62}$$

Since (57) is held, and the small term $\delta\left(\mathbf{Q}_j(t), \Lambda_j \mathbf{d}_j(t)\right) > 0$ when Λ_j is sufficiently small, the right side of the equation (62) is larger than 0. Noting $h_i(\mathbf{Q}_j(t)) = 0$ for $i \in A(\mathbf{Q}_j(t))$ we can see the left side of the equation (62) is also larger than 0, i.e., $h_i(\mathbf{Q}_j(t) + \Lambda_j \mathbf{d}_j(t)) > 0$.

To sum up, $h_i(\mathbf{Q}_j(t) + \Lambda_j \mathbf{d}_j(t)) \geq 0$ can be held for all i when a certain Λ_j is sufficiently small. This implies that once the step size Λ_j is well bounded, a new iterator generated by the iteration direction $\mathbf{d}_j(t)$ satisfying the conditions above, $\mathbf{Q}_j(t) + \Lambda_j \mathbf{d}_j(t)$, is alway feasible for (18). Additionally, recalling the definition of $\mathbf{G}_j\left(\mathbf{Q}_j(t), \mathbf{D}_j(t)\right)$, the condition (56) confirms that the value of the objective function in (18) can be improved at the iteration direction $\mathbf{d}_j(t)$. At this point, this lemma is proven. ■

B. Lemma 2

Lemma 2. *Suppose that $\mathbf{Q}_j(t)$ is a feasible point at the t-th iteration and an optimal value of the objective function of (41) obtained at this iterator $\mathbf{Q}_j(t)$ is ω. If and only if ω satisfies $\omega = 0$, then $\mathbf{Q}_j(t)$ is a Fritz John point.*

Proof: Recalling (41), only the constraints (41b) and (41c) involve the parameter ω. Indeed, ω can be regarded as a type of slack variable in these constraints. Obviously, $\omega = 0$ indicates that the system $\mathbf{G}_j\left(\mathbf{Q}_j(t), \mathbf{D}_j(t)\right) \mathrm{vec}(\mathbf{d}_j(t)) > 0$ and $\bigtriangledown_{\mathbf{Q}_j} h_i\left(\mathbf{Q}_j(t)\right) \mathrm{vec}(\mathbf{d}_j(t)) > 0$ for $i \in A(\mathbf{Q}_j(t)) := \{i \,|\, h_i(\mathbf{Q}_j(t)) = 0, i \in \mathscr{J}\}$ has no solutions. Otherwise, if this system had a solution, there would exist a negative real number which is near zero, denoted by $\bar{\omega} < \omega = 0$, such that $\mathbf{G}_j\left(\mathbf{Q}_j(t), \mathbf{D}_j(t)\right) \mathrm{vec}(\mathbf{d}_j(t)) + \bar{\omega} \geq 0$ and $\bigtriangledown_{\mathbf{Q}_j} h_i\left(\mathbf{Q}_j(t)\right) \mathrm{vec}(\mathbf{d}_j(t)) + \bar{\omega} \geq 0$ for $i \in A(\mathbf{Q}_j(t))$. In this case, $\omega = 0$ would violate its optimality (namely, ω is not an optimal function value, which is contradictory to the given condition). Furthermore, according to the Gordan theorem, a necessary and sufficient condition of the inequality system having no solutions is given as follows:

There exist some certain coefficients, $\{v_0 \geq 0, v_i \geq 0, i \in A(\mathbf{Q}_j(t))\}$, that are not all to be zero and that can make the following equation always held

$$v_0 \mathbf{G}_j\left(\mathbf{Q}_j(t), \mathbf{D}_j(t)\right) + \sum_{i \in A(\mathbf{Q}_j(t))} v_i \bigtriangledown_{\mathbf{Q}_j} h_i\left(\mathbf{Q}_j(t)\right) = 0 \tag{63}$$

In fact, according to the definition of a Fritz John point of an optimization problem, this argument mentioned above indicates that $\mathbf{Q}_j(t)$ is a Fritz John point. Hence, this proof is complete. ■

C. The proof of Corollary 1

Firstly, we expand γ_j as

$$\begin{aligned} \gamma_j =& \mathrm{Tr}\left(\widetilde{\mathbf{U}}_j \mathbf{Q}_j \widetilde{\mathbf{U}}_j^{\dagger}\right) + \mathrm{Tr}\left(\widetilde{\mathbf{U}}_j \mathbf{Q}_j \Delta \mathbf{U}_j^{\dagger}\right) \\ &+ \mathrm{Tr}\left(\Delta \mathbf{U}_j \mathbf{Q}_j \widetilde{\mathbf{U}}_j^{\dagger}\right) + \mathrm{Tr}\left(\Delta \mathbf{U}_j \mathbf{Q}_j \Delta \mathbf{U}_j^{\dagger}\right). \end{aligned} \tag{64}$$

Since $\mathbf{Q}_j \in \mathbb{M}_+^{m_j \times m_j}$, $\mathbf{Q}_j = \mathbf{Q}_j^{\dagger}$ always holds. We rearrange

$$\begin{aligned} &\mathrm{Tr}\left(\widetilde{\mathbf{U}}_j \mathbf{Q}_j \Delta \mathbf{U}_j^{\dagger}\right) + \mathrm{Tr}\left(\Delta \mathbf{U}_j \mathbf{Q}_j \widetilde{\mathbf{U}}_j^{\dagger}\right) \\ &= 2\mathrm{vec}\left(\mathbf{Q}_j^{\dagger} \widetilde{\mathbf{U}}_j^{\dagger}\right)^{\dagger} \mathrm{vec}\left(\Delta \mathbf{U}_j^{\dagger}\right). \end{aligned} \tag{65}$$

Similarly, we can directly transform $\mathrm{Tr}\left(\Delta \mathbf{U}_j \mathbf{Q}_j \Delta \mathbf{U}_j^{\dagger}\right)$ to

$$\mathrm{Tr}\left(\Delta \mathbf{U}_j \mathbf{Q}_j \Delta \mathbf{U}_j^{\dagger}\right) = \left(\mathrm{vec}\left(\Delta \mathbf{U}_j^{\dagger}\right)\right)^{\dagger} \left(\mathbf{I}_u \otimes \mathbf{Q}_j\right) \mathrm{vec}\left(\Delta \mathbf{U}_j^{\dagger}\right) \tag{66}$$

Moreover, combining (64), (65) and (66), and substituting the expansion of γ_j into the inequality $\gamma_j^{\max}-\gamma_j \geq 0$ immediately yields the equivalent inequality (16) in this Corollary. In addition, it is obvious that the constraint associated with $\Delta\mathbf{U}_j$, i.e., $\mathrm{Tr}\left(\Delta\mathbf{U}_j\Delta\mathbf{U}_j^{\dagger}\right) \leq \sigma_j^2$, can be reformulated as $\|\Delta\mathbf{U}_j\|_2 \leq \sigma_j$. Additionally, according to the S-procedure in [27], the inequality (16) can be easily re-expressed by a linear matrix inequality (17). At this point, Corollary 1 is proven.

D. The proof of Corollary 2

By contradiction, we assume that a certain subsequence $\{(\mathbf{Q}_j(t), \mathbf{d}_j(t)), t \in \mathscr{T}\}$ satisfying these four conditions exists. Note that the objective function g_j is continuously differentiable. Since $\mathbf{d}_j(t) \to \mathbf{d}_j$ and $\mathbf{Q}_j(t) \to \mathbf{Q}_j$ for $t \in \mathscr{T}$, and the condition (iv), there must exist a positive real number $a_1 > 0$ such that $\mathbf{G}_j\left(\mathbf{Q}_j(t), \mathbf{D}_j(t)\right)\mathrm{vec}\left(\mathbf{d}_j(t)\right) = a_1 > 0$ for $t \in \mathscr{T}$ sufficiently large.

Additionally, there also exists a $a_2 > 0$ such that $\mathbf{G}_j\left(\mathbf{Q}_j(t) + \Lambda_j'\mathbf{d}_j(t), \mathbf{D}_j(t)\right)\mathrm{vec}\left(\mathbf{d}_j(t)\right) = a_2 > 0$ for $t \in \mathscr{T}$ sufficiently large. According to the intermediate value theorem, for any $\widetilde{\Lambda}_j \in \left(0, \Lambda_j'\right)$ there must exist such a corresponding positive real number $\tilde{a} \in (\min\{a_1, a_2\}, \max\{a_1, a_2\})$ that satisfies

$$\mathbf{G}_j\left(\mathbf{Q}_j(t) + \widetilde{\Lambda}_j\mathbf{d}_j(t), \mathbf{D}_j(t)\right)\mathrm{vec}\left(\mathbf{d}_j(t)\right) = \tilde{a} > 0 \tag{67}$$

for $t \in \mathscr{T}$ sufficiently large.

By (iii) and by the definition of $\mathbf{Q}_j(t+1)$, we have

$$\begin{aligned} &g_j\left(\mathbf{Q}_j(t+1), \mathbf{D}_j(t) \,|\mathscr{H}_j, \mathscr{Q}_{-j}(t)\right) \\ &\geq g_j\left(\mathbf{Q}_j(t) + \widetilde{\Lambda}_j'\mathbf{d}_j(t), \mathbf{D}_j(t) \,|\mathscr{H}_j, \mathscr{Q}_{-j}(t)\right) \end{aligned} \tag{68}$$

where $0 < \widetilde{\Lambda}_j' \leq \Lambda_j'$.

Using the mean value theorem, we can expand the right term of the inequality above as follows

$$\begin{aligned} &g_j\left(\mathbf{Q}_j(t) + \widetilde{\Lambda}_j'\mathbf{d}_j(t), \mathbf{D}_j(t) \,|\mathscr{H}_j, \mathscr{Q}_{-j}(t)\right) \\ &= g_j\left(\mathbf{Q}_j(t), \mathbf{D}_j(t) \,|\mathscr{H}_j, \mathscr{Q}_{-j}(t)\right) \\ &\quad + \widetilde{\Lambda}_j'\mathbf{G}_j\left(\widetilde{\mathbf{Q}}_j(t), \mathbf{D}_j(t)\right)\mathrm{vec}\left(\mathbf{d}_j(t)\right) \end{aligned} \tag{69}$$

where $\widetilde{\mathbf{Q}}_j(t) = \mathbf{Q}_j(t) + \xi\widetilde{\Lambda}_j'\mathbf{d}_j(t)$ and $\xi \in (0,1)$.

Then, by setting $\widetilde{\Lambda}_j = \xi\widetilde{\Lambda}_j'$ and applying (67) we can yield $\mathbf{G}_j\left(\widetilde{\mathbf{Q}}_j(t), \mathbf{D}_j(t)\right)\mathrm{vec}\left(\mathbf{d}_j(t)\right) = \tilde{a}$. Furthermore, combining this result with (68) and (69) immediately arrives at

$$\begin{aligned} &g_j\left(\mathbf{Q}_j(t+1), \mathbf{D}_j(t) \,|\mathscr{H}_j, \mathscr{Q}_{-j}(t)\right) \\ &\geq g_j\left(\mathbf{Q}_j(t), \mathbf{D}_j(t) \,|\mathscr{H}_j, \mathscr{Q}_{-j}(t)\right) + \widetilde{\Lambda}_j'\tilde{a} \end{aligned} \tag{70}$$

for $t \in \mathscr{T}$ sufficiently large.

Note that iterators $\{\mathbf{Q}_j(t)\}$ generated by the corresponding feasible direction sequence $\{\mathbf{d}_j(t)\}$ sequentially improves value of the objective function. Using (ii) we can get

$$\begin{aligned} &\lim_{t\to+\infty} g_j\left(\mathbf{Q}_j(t+1), \mathbf{D}_j(t) \,|\mathscr{H}_j, \mathscr{Q}_{-j}(t)\right) \\ &= \lim_{t\to+\infty} g_j\left(\mathbf{Q}_j(t), \mathbf{D}_j(t) \,|\mathscr{H}_j, \mathscr{Q}_{-j}(t)\right) \\ &= g_j\left(\mathbf{Q}_j, \mathbf{D}_j(t) \,|\mathscr{H}_j, \mathscr{Q}_{-j}(t)\right) \end{aligned} \tag{71}$$

This implies that when $t \in \mathscr{T}$ approaches the positive infinity (70) will result in

$$\begin{aligned} &g_j\left(\mathbf{Q}_j, \mathbf{D}_j(t) \,|\mathscr{H}_j, \mathscr{Q}_{-j}(t)\right) \\ &\geq g_j\left(\mathbf{Q}_j, \mathbf{D}_j(t) \,|\mathscr{H}_j, \mathscr{Q}_{-j}(t)\right) + \widetilde{\Lambda}_j'\tilde{a} \\ &> g_j\left(\mathbf{Q}_j, \mathbf{D}_j(t) \,|\mathscr{H}_j, \mathscr{Q}_{-j}(t)\right) \end{aligned} \tag{72}$$

Thus, the contradiction occurs, and the Corollary 2 is proven.

REFERENCES

[1] E. Hossain, G. Chow, V. C. Leung, R. D. McLeod, J. Misic, V. W. Wong, and O. Yang, "Vehicular telematics over heterogeneous wireless networks: A survey," *Computer Communications*, vol. 33, no. 7, pp. 775 – 793, 2010.

[2] N. Lu, N. Cheng, N. Zhang, X. Shen, and J. Mark, "Connected vehicles: Solutions and challenges," *Internet of Things Journal, IEEE*, vol. 1, no. 4, pp. 289–299, Aug 2014.

[3] J. Wang, D. Ni, and K. Li, "Rfid-based vehicle positioning and its applications in connected vehicles," *Sensors*, vol. 14, no. 3, pp. 4225–4238., 2013.

[4] F. Qu, F. Wang, and L. Yang, "Intelligent transportation spaces: vehicles, traffic, communications, and beyond," *IEEE Communications Magazine*, vol. 48, no. 11, pp. 136–142, 2010.

[5] M. Chen, Y. Hao, Y. Li, C. Lai, and D. Wu, "On the computation offloading at ad hoc cloudlet: Architecture and service models," *IEEE Communications*, vol. 53, no. 6, pp. 18–24, 2015.

[6] M. Chen, Y. Zhang, Y. Li, S. Mao, and V. C. Leung, "Emc: Emotion-aware mobile cloud computing in 5g," *IEEE Network*, vol. 29, no. 2, pp. 32–38, 2015.

[7] M. Pan, P. Li, and Y. Fang, "Cooperative communication aware link scheduling for cognitive vehicular networks," *Selected Areas in Communications, IEEE Journal on*, vol. 30, no. 4, pp. 760–768, May 2012.

[8] A. Goldsmith, S. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of mimo channels," *Selected Areas in Communications, IEEE Journal on*, vol. 21, no. 5, pp. 684–702, June 2003.

[9] F. Qu, Z. Zhang, Z. Wu, and Y. Wu, "Robust transceiver optimization for multi-antenna multi-carrier multi-hop communications," *China Communications*, vol. 13, no. 4, pp. 20–29, 2016.

[10] M. Ismail and W. Zhuang, "Network cooperation for energy saving in green radio communications," *Wireless Communications, IEEE*, vol. 18, no. 5, pp. 76–81, October 2011.

[11] A. Hammad, T. Todd, G. Karakostas, and D. Zhao, "Downlink traffic scheduling in green vehicular roadside infrastructure," *Vehicular Technology, IEEE Transactions on*, vol. 62, no. 3, pp. 1289–1302, March 2013.

[12] C. Yang, Y. Fu, Y. Zhang, S. Xie, and R. Yu, "Energy-efficient hybrid spectrum access scheme in cognitive vehicular ad hoc networks," *Communications Letters, IEEE*, vol. 17, no. 2, pp. 329–332, February 2013.

[13] R. Chen, Z. Sheng, Z. Zhong, M. Ni, V. C. Leung, D. Michelson, and M. Hu, "Connectivity analysis for cooperative vehicular ad hoc networks under nakagami fading channel," *IEEE Communications Letters*, vol. 18, no. 10, pp. 1787 – 1790, 2014.

[14] X. Cheng, X. Hu, L. Yang, I. Husain, K. Inoue, P. Krein, R. Lefevre, Y. Li, H. Nishi, J. Taiber, F.-Y. Wang, Y. Zha, W. Gao, and Z. Li, "Electrified vehicles and the smart grid: The its perspective," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 15, no. 4, pp. 1388–1404, Aug 2014.

[15] H. Kwon and T. Birdsall, "Channel capacity in bits per joule," *Oceanic Engineering, IEEE Journal of*, vol. 11, no. 1, pp. 97–99, Jan 1986.

[16] Y.-C. Liang, K.-C. Chen, G. Li, and P. Mahonen, "Cognitive radio networking and communications: an overview," *Vehicular Technology, IEEE Transactions on*, vol. 60, no. 7, pp. 3386–3407, Sept 2011.

[17] T.-D. Nguyen, O. Berder, and O. Sentieys, "Energy-efficient cooperative techniques for infrastructure-to-vehicle communications," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 12, no. 3, pp. 659–668, Sept 2011.

[18] J. Xu and L. Qiu, "Energy efficiency optimization for mimo broadcast channels," *Wireless Communications, IEEE Transactions on*, vol. 12, no. 2, pp. 690–701, February 2013.

[19] C. Jiang and L. Cimini, "Energy-efficient transmission for mimo interference channels," *Wireless Communications, IEEE Transactions on*, vol. 12, no. 6, pp. 2988–2999, June 2013.

[20] W. Zhong and J. Wang, "Energy efficient spectrum sharing strategy selection for cognitive mimo interference channels," *Signal Processing, IEEE Transactions on*, vol. 61, no. 14, pp. 3705–3717, July 2013.

[21] X. Zhang, H. Li, Y. Lu, and B. Zhou, "Distributed energy efficiency optimization for mimo cognitive radio network," *IEEE Communications Letters*, vol. 19, pp. 847–850, 2015.

[22] G. Zheng, K.-K. Wong, and B. Ottersten, "Robust cognitive beamforming with bounded channel uncertainties," *Signal Processing, IEEE Transactions on*, vol. 57, no. 12, pp. 4871–4881, Dec 2009.

[23] G. Zheng, S. Ma, K.-K. Wong, and T.-S. Ng, "Robust beamforming in cognitive radio," *Wireless Communications, IEEE Transactions on*, vol. 9, no. 2, pp. 570–576, February 2010.

[24] J. Wang, G. Scutari, and D. Palomar, "Robust mimo cognitive radio via game theory," *Signal Processing, IEEE Transactions on*, vol. 59, no. 3, pp. 1183–1201, March 2011.

[25] H. Tembine, "Dynamic robust games in mimo systems," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 41, no. 4, pp. 990–1002, Aug 2011.

[26] Y. Zhang, E. DallAnese, and G. Giannakis, "Distributed optimal beamformers for cognitive radios robust to channel uncertainties," *Signal Processing, IEEE Transactions on*, vol. 60, no. 12, pp. 6495–6508, Dec 2012.

[27] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge University Press, 2004.

[28] L. Fu, Y. Zhang, and J. Huang, "Energy efficient transmissions in mimo cognitive radio networks," *Selected Areas in Communications, IEEE Journal on*, vol. 31, no. 11, pp. 2420–2431, November 2013.

[29] J. Mao, G. Xie, J. Gao, and Y. Liu, "Energy efficiency optimization for cognitive radio mimo broadcast channels," *Communications Letters, IEEE*, vol. 17, no. 2, pp. 337–340, February 2013.

[30] L. Zhang, Y. Xin, and Y.-C. Liang, "Weighted sum rate optimization for cognitive radio mimo broadcast channels," in *Communications, 2008. ICC '08. IEEE International Conference on*, May 2008, pp. 3679–3683.

[31] S. S. Haykin, "Adaptive filter theory," *Adaptive Filter Theory*, vol. 4, no. 96, 2002.

[32] A. Hjrungnes, *Complex-Valued Matrix Derivatives*. Cambridge University Press, 2011.

[33] W. Dinkelbach, "On nonlinear fractional programming," *Management Science*, vol. 13, no. 7, pp. 492–498, 1967.

[34] J. Nocedal and S. J. Wright, *Numerical Optimization, second edition*. World Scientific, 2006.

[35] J. Lofberg, "Yalmip: A toolbox for modeling and optimization in MATLAB," in *Proceedings of the CACSD Conference*, Taipei, Taiwan, 2004.

[36] J. F. Sturm, "Implementation of interior point methods for mixed semidefinite and second order cone optimization problems," *Optimization Methods & Software*, vol. 17, no. 6, pp. 1105–1154, 2002.

[37] A. Molisch, F. Tufvesson, J. Karedal, and C. Mecklenbrauker, "A survey on vehicle-to-vehicle propagation channels," *Wireless Communications, IEEE*, vol. 16, no. 6, pp. 12–22, December 2009.

[38] L. Cheng, B. Henty, D. Stancil, F. Bai, and P. Mudalige, "Mobile vehicle-to-vehicle narrow-band channel measurement and characterization of the 5.9 ghz dedicated short range communication (dsrc) frequency band," *Selected Areas in Communications, IEEE Journal on*, vol. 25, no. 8, pp. 1501–1516, Oct 2007.

[39] A. Akki and F. Haber, "A statistical model of mobile-to-mobile land communication channel," *Vehicular Technology, IEEE Transactions on*, vol. 35, no. 1, pp. 2–7, Feb 1986.

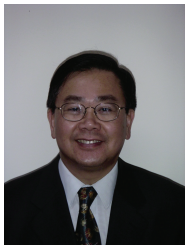
[40] C. S. Patel, G. L. Stuber, and T. G. Pratt, "Simulation of rayleigh-faded mobile-to-mobile communication channels," *Communications, IEEE Transactions on*, vol. 53, no. 10, pp. 1773–1773, Oct 2005.

[41] J. Xu, X. Yu, W. Lu, F. Qu, and N. Dengf, "Offset manchester coding for rayleigh noise suppression in carrier-distributed wdm-pons," *Optics Communications*, vol. 346, pp. 106–109, 2015.

Daxin Tian [M'13, SM'16] is an associate professor in the School of Transportation Science and Engineering, Beihang University, Beijing, China. His current research interests include mobile computing, intelligent transportation systems, vehicular ad hoc networks, and swarm intelligent.

Jianshan Zhou received the B.Sc. and M.Sc. degrees in traffic information engineering and control in 2013 and 2016, respectively. He is currently working toward the Ph.D degree with the School of Transportation Science and Engineering, Beihang University, Beijing, China. His current research interests are focused on wireless communication, artificial intelligent system, and intelligent transportation systems.

Zhengguo Sheng is currently a Lecturer with the Department of Engineering and Design, University of Sussex, U.K. He has authored over 50 international conference and journal papers. His current research interests cover IoT/M2M, vehicular communications, and edge/cloud computing.



Victor C. M. Leung [S'75, M'89, SM'97, F'03] is a Professor of Electrical and Computer Engineering and holder of the TELUS Mobility Research Chair at the University of British Columbia (UBC). His research is in the areas of wireless networks and mobile systems. He has co-authored more than 900 technical papers in archival journals and refereed conference proceedings, several of which had won best-paper awards. Dr. Leung is a Fellow of the Royal Society of Canada, a Fellow of the Canadian Academy of Engineering and a Fellow of the Engineering Institute of Canada. He is serving on the editorial boards of IEEE JSAC-SGCN, IEEE Wireless Communications Letters, IEEE Access and several other journals. He has provided leadership to the technical program committees and organizing committees of numerous international conferences. Dr. Leung was the recipient of the 1977 APEBC Gold Medal, NSERC Postgraduate Scholarships from 1977-1981, a 2012 UBC Killam Research Prize, and an IEEE Vancouver Section Centennial Award.